The convergent evolution of caste in ants and honey bees is based on a shared core of ancient reproductive genes and many plastic genes

Michael R. Warner¹, Lijun Qiu², Michael J. Holmes²-³, Alexander S. Mikheyev²-⁴, Timothy A. Linksvayer¹

¹University of Pennsylvania, Philadelphia, Pennsylvania, USA
²Okinawa Institute of Science and Technology, Okinawa, Japan
³School of Life and Environmental Science, University of Sydney, Australia
⁴Research School of Biology, Australian National University, Canberra, Australia

Abstract

Eusociality, characterized by caste-based division of labor, has convergently evolved multiple times. However, the genomic basis of caste and degree to which independent origins of eusociality have utilized common genes is largely unknown. To elucidate these issues, we characterized caste-specific transcriptomic profiles across development and adult body segments from honey bees (Apis mellifera) and pharaoh ants (Monomorium pharaonis), representing two independent origins of eusociality. We identified a shared core of genes upregulated in the abdomens of queen honey bees and ants that is also upregulated in female flies. Outside of this shared core, few genes are differentially expressed in common. Instead, the majority of genes underlying the caste system show plastic expression, are rapidly evolving, and are relatively evolutionary young. Altogether our results show that the convergent evolution of eusociality involves the recruitment of a core reproductive groundplan along with many plastically-expressed and rapidly evolving genes.
Introduction

Elucidating the genetic basis and evolution of phenotypic novelty is a major goal of evolutionary biology (Wray et al. 2003; Muller & Wagner 1991; Beldade & Brakefield 2002; Carroll 2008; Hoekstra & Coyne 2007). The field of evolutionary developmental biology (“evo-devo”) has demonstrated that cis-regulatory mutations in a small set of highly conserved genes -- often described as a “genetic toolkit” -- play a major role in generating morphological novelty and diversity by altering body plan formation early in development (Carroll 2008; Hoekstra & Coyne 2007; Wray 2007; Prud’homme et al. 2006; Wittkopp & Kalay 2011; Davidson 2010).

The importance of such a genetic toolkit is seen clearly in examples of convergent phenotypic evolution. For example, eye development in both vertebrates and arthropods is governed by the transcription factor *Pax6* despite the fact that these lineages evolved eyes independently (Halder et al. 1995), and likewise vertebrate and arthropod appendage evolution has repeatedly utilized copies of the gene *Distal-less* to produce morphological novelty (Nielsen & Martinez 2003; Shubin et al. 2009). Furthermore, this process often involves the co-option of general pathways not directly related to the novel trait, as is the case in the utilization of the conserved sex-determination gene *doublesex* in novel polyphenisms in insects (Kunte et al. 2014; Kijimoto et al. 2012).

At the same time, an unexpected result of comparative genomics research has been the widespread occurrence of genes that are unique to a species or lineage (Wilson et al. 2005; Tautz & Domazet-Lošo 2011; Wilson et al. 2007), and these taxonomically-restricted genes also appear to play important roles in the evolution of phenotypic novelty (Khalturin et al. 2009; Jasper et al. 2015; Chen et al. 2013; Wilson et al. 2005). For example, taxonomically-restricted genes have
been associated with the evolution of novel cell types and organs, such as cnidarian nematocysts
(Khalturin et al. 2008; Babonis et al. 2016; Milde et al. 2009; Hwang et al. 2007), water strider
propeller fans (Santos et al. 2017), and the molluscan radula (Hilgers et al. 2018).

As a result of these different findings, some researchers stress the importance of highly
conserved genes (Carroll 2008; Müller 2007; Wagner 2014), while others stress the importance
of taxonomically-restricted genes for the evolution of phenotypic novelty (Khalturin et al. 2009;
Jasper et al. 2015; Chen et al. 2013; Wilson et al. 2005). However, the relative importance of
each type of gene to the full genetic architecture underlying phenotypic novelty is unresolved
(Johnson 2018). Furthermore, it remains unclear whether all types of novel traits (e.g.,
morphological, physiological, behavioral) have similar types of genetic and evolutionary features
(Mikheyev & Linksvayer 2015; Jasper et al. 2015).

Some of the most striking phenotypic innovations involve the evolution of novel social
behavior and social complexity. This is exemplified by the caste-based division of labor that has
convergently evolved in several eusocial insect lineages (e.g., ants, honey bees, vespid wasps,
termites) (Johnson & Linksvayer 2010; Hölldobler & Wilson 2009). So far, sociogenomic
research has proceeded along generally similar lines to research on the genetic basis and
evolution of morphological novelty, with researchers alternately emphasizing conserved versus
taxonomically-restricted genes. Following findings from evo-devo and the study of
morphological innovation, researchers proposed that the evolution of novel social behavior is
similarly driven by changing the regulation of a core set of genes, which were hypothesized to
comprise a “genetic toolkit” for the evolution of social behavior (Toth & Robinson 2007; Toth et
al. 2010; Rittschof & Robinson 2016; O’Connell & Hofmann 2012; Saul et al. 2018). This was
largely motivated by the observation that highly conserved genes and gene pathways including insulin signaling, vitellogenin, and juvenile hormone regulate many aspects of complex societies, most prominently reproductive caste development (Wheeler 1986; Trible & Kronauer 2017; Libbrecht et al. 2013; Wheeler et al. 2006; Okada et al. 2017) and worker division of labor (Smith et al. 2008; Robinson 1987; Sullivan et al. 2000; Amdam et al. 2004; Page & Amdam 2007). Alternatively, many other recent studies have emphasized the role of taxonomically-restricted genes in the evolution of social complexity, in particular associated with the worker caste (Feldmeyer et al. 2014; Warner et al. 2017; Ferreira et al. 2013; Jasper et al. 2015; Johnson & Tsutsui 2011; Harpur et al. 2014; Mikheyev & Linksvayer 2015; Patalano et al. 2015) and tissues associated with social communication (Jasper et al. 2015). Altogether, the genomic basis of caste-based division of labor remains largely unclear, as does the degree to which independent origins of eusociality have utilized common genes.

Here, we present to date the most comprehensive developmental transcriptomic dataset investigating gene expression associated with reproductive dimorphism between queens and workers and age-based division of labor between worker nurses and foragers in the honey bee (Apis mellifera) and the pharaoh ant (Monomorium pharaonis), species which represent two independent origins and elaborations of eusociality (Branstetter et al. 2017). We performed all sampling, sequencing, and analysis for the two species in parallel to maximize compatibility between the data sets. We leverage the separate origins of eusociality to assess in an unbiased manner the relative importance of shared versus lineage-specific genetic mechanisms to the convergent evolution of reproductive caste and worker division of labor as well as the relative importance of ancient versus taxonomically-restricted genes at each life stage and tissue to the
expression of these traits. We identify a core set of reproduction-related genes upregulated in
honey bee and ant queen abdomens, which is derived from a set of ancient sex-biased genes.

However, this core set comprises only one third of abdominal differential expression, and outside
the abdomen, lineage-specific changes predominate, where young, loosely connected, and
plastically expressed genes tend to underlie division of labor.

Results

We constructed two large, parallel transcriptomic datasets in honey bees and pharaoh ants
spanning caste development as well as adult tissues separated by behavior, reproductive caste,
and sex. In total sequenced 177 total RNA sequencing libraries across 28 distinct sample types
(Table S1).

Differential expression between queens and workers

To identify genes associated with caste development and caste dimorphism, we performed
differential expression analysis between queens and workers at each stage and adult body
segment, separately for each species. The number of differentially expressed genes (DEGs)
between queens and workers at a given developmental stage or tissue varied between 245 in
honey bee pupae and 5352 in honey bee abdomens (Table S2) and generally increased
throughout development. The largest transcriptomic signature of caste occurred in the adult
abdomen (Figure 1A). In all tissues and stages, the majority of DEGs between queens and
workers in one species were not differentially expressed in the other species (Figure 1A).

Similarly, the magnitude of gene-wise caste bias (as measured by log2 fold-change between
queen and worker samples) was positively correlated between ant and honey bee orthologs in all three adult tissues, with the strongest correlation in the abdomen, but uncorrelated or negatively correlated in all larval and pupal stages (Figure S1; $r_{\text{head}} = 0.089; r_{\text{thorax}} = 0.161; r_{\text{abdomen}} = 0.275$; $N = 7640$ one-to-one orthologs; $N = 7460$ one-to-one orthologs; $P < 0.001$ in all cases). The top enriched Gene Ontology (GO) terms for DEGs between queens and workers were dominated by metabolism, signaling and developmental processes (Table S3,S4).

**Differential expression between nurses and foragers**

To identify genes associated with age-based worker division of labor, we performed differential expression analysis between nurses and foragers in adult body segment, separately for each species. The number of DEGs between nurses and foragers at a given tissue varied from 405 in ant heads to 2519 in honey bee thoraces (Table S5). There were more DEGs between nurses and foragers in honey bees than in ants for all three adult tissues. In general there were very few DEGs in common in the two species between nurses and foragers (Figure 1B). Gene-wise log$_2$ fold-change between nurses and foragers was significantly but weakly correlated across ant and honey bee orthologs (Figure S2; $r_{\text{head}} = 0.070$, $P_{\text{head}} < 0.001$; $r_{\text{thorax}} = 0.031$, $P_{\text{thorax}} = 0.008$; $r_{\text{abdomen}} = 0.051$, $P_{\text{abdomen}} < 0.001$; $N = 7460$ one-to-one orthologs). The top enriched GO terms based for DEGs between nurses and foragers were dominated by metabolism and developmental processes (Table S6,S7).

**Conserved abdominal caste-bias in ancient genes**
For the most part our results indicate distinct genes are associated with caste and worker division of labor in honey bees and ants. However, a third of DEGs between queen and worker abdomens were common to both species (Figure 1A; 1545 shared DEGs in abdomen, comprising 35% [1545/4395] of ant abdominal DEGs, and 29% [1545/5352] of honey bee abdominal DEGs). Most shared abdominal differential expression was the result of shared queen-bias: 56% (858/1545 genes) of commonly caste-biased genes were upregulated in queen abdomens in both species, compared to 22% (338/1545) that were worker-upregulated and 23% (349/1545) that reversed direction (i.e. were queen-biased in one species and worker-biased in the other). Genes with conserved abdominal queen-bias tended to be evolutionarily older than genes with conserved worker bias or non-conserved bias (Figure 1C), though the estimated evolutionary age of genes was not consistently associated with queen or worker bias across all tissues and stages (Figure S3).

To identify core genes associated with queen abdominal expression, we performed gene coexpression analysis, separately for each species. We focused on identifying a group of genes (i.e. module) which exhibited correlated expression patterns specifically across queen abdominal samples (see Methods). We identified a queen abdominal module of 1006 genes in ants and a module of 1174 genes in honey bees. We calculated the connectivity of each gene within these identified abdominal modules (i.e. intra-modular connectivity), where genes with higher connectivity values are central to the module, i.e. they exhibit highly correlated expression patterns with many other genes in the module. There was a positive correlation between intra-modular connectivity and log2 fold-change between queen and worker abdomens (Figure 2A,B; ants: rho = 0.536, P < 0.001, N = 1006; honey bees: rho = 0.617, P < 0.001, N = 1174), as well a
positive correlation between intra-modular connectivity and the significance (-log₁₀ of P-value) of queen/worker abdominal differential expression (Spearman correlation; ants: rho = 0.485, P < 0.001, N = 1006; honey bees: rho = 0.555, P < 0.001, N = 1174). We identified 183 genes which were present in the queen abdominal module of both species, and these genes were overwhelmingly queen-biased (78.7% [144/183] upregulated in queens of both species), and had higher intra-module connectivity than genes found in only one species-specific module (Figure 2C,D).

A major strength of gene coexpression analysis is its power to identify hub genes for traits, i.e. genes which are centrally connected and strongly associated with a trait (Zhang & Horvath 2005). We identified hub genes in both species (see Methods), many of which are clearly associated with reproduction and maternal effects (Table S8,S9; Figure 2A,B). Included in this list are genes with previous known roles in caste and reproduction such as vitellogenin (Vg receptor was identified in each species) (Barchuk et al. 2002; Libbrecht et al. 2013) and vasa (Khila & Abouheif 2010), while others are important maternal proteins such as Smaug (Benoit et al. 2009) and ovo (Mével-Ninio et al. 1995).

Caste bias is in part derived from ancestral sex bias

Given that our coexpression analysis indicated that many of the most important queen-associated genes are clearly associated with female reproduction, we reasoned that caste-biased expression may be derived from sex-biased expression. Indeed, there was a positive correlation between gene-wise log₂ fold change between queen and worker abdomens and gene-wise log₂ fold-change between queen and male abdomens in both honey bees and pharaoh ants (Figure 3A,B).
Additionally, sex bias itself was correlated between species (Figure 3C). The correlation of caste- and sex-bias was not restricted to the abdomen, as there was similar highly significant effect when comparing heads and thoraces, albeit with weaker effect sizes (Figure S4).

Given the link between conserved caste bias and sex bias within ants and honey bees, we hypothesized that caste bias is derived from ancient pathways underlying sexual dimorphism. To test this hypothesis, we estimated the whole-body sex-bias of orthologs in the fruitfly *Drosophila melanogaster* using available data (Gerstein et al. 2014). Genes with consistent queen-bias in our ant and honey bee abdomen samples tended to be upregulated in females in *D. melanogaster* (Figure 3D; one-sided Binomial Test for likelihood of queen conserved having log_2 fold-change > 0; P < 0.001; N = 566 conserved queen genes), while genes with consistent worker-bias tended to be upregulated in males (P < 0.001; N = 160 conserved worker genes).

*Expression plasticity across development, caste, and tissue is correlated between species*

While we have emphasized the conservation of abdominal differential expression between queens and workers in pharaoh ants and honey bees, roughly two-thirds of abdominal caste-biased genes were not shared between species, and differential expression based on reproductive caste or worker division of labor was largely not shared (Figure 1A,B). Furthermore, genes were often differentially expressed across many stages and tissues and sometimes in opposite directions (e.g., upregulated in queen heads but downregulated in queen abdomens), while other genes showed little to no expression differences (Figure S5). To quantify the degree to which genes exhibited biased expression according to reproductive caste across all developmental stages and tissues we calculated the Euclidean distance of log_2 fold-change across all
queen/worker comparisons separately for each species, and we labeled this quantity “overall caste bias” (this approach is analogous to previous work on expression plasticity in ants across alternative morphs, see (Schrader et al. 2017)). Similarly, we defined “overall behavior bias” as the Euclidean distance of log₂ fold-change across all nurse/forager comparisons, separately for each species.

Interestingly, both overall caste and overall behavior bias were correlated among orthologs between species (Figure S6; i.e. caste bias measured in ants was correlated with caste bias measured in honey bees across 1-1 orthologs). Within species, caste and behavior bias were correlated to each other (Figure S7). This indicates that plasticity in gene expression is correlated across contexts (caste versus behavior) and species. GO terms associated with high caste bias were largely linked to metabolism, while those associated with high behavior bias were largely linked to developmental processes (Table S10).

**Characteristics of genes associated with caste and behavior**

We compared overall caste bias and overall behavior bias to gene age, evolutionary rate, network connectivity, and tissue-specificity to understand the general features of genes commonly associated with caste (queen versus worker) or behavior (nursing versus foraging). Genes with younger estimated evolutionary ages tended to exhibit higher overall caste bias (Figure 4A,B) and behavior bias (Figure S8A,B) compared in particular to ancient genes (Gamma GLM; ant caste bias: $\chi^2 = 900.19$, honey bee caste bias: $\chi^2 = 1412.80$, ant behavior bias: $\chi^2 = 316.36$, honey bee behavior bias: $\chi^2 = 877.43$; P < 0.001 for all cases; N = 10520 in ant, N = 10011 in honey bees). Genes that were loosely connected in coexpression networks constructed across all
samples tended to exhibit more caste and behavior bias in comparison to highly connected genes (Figure 4C,D, Figure S8C,D). Genes with low connectivity are peripheral network elements, while genes with high connectivity are central to regular networks and often highly pleiotropic (Zhang & Horvath 2005). Similarly, genes with high tissue-specificity across 12 honey bee tissues tended to exhibit higher values caste and behavior bias in comparison to more pleiotropic, ubiquitously expressed genes (Figure S9), where tissue specificity was calculated using available data (Jasper et al. 2015). Finally, genes that were rapidly evolving (i.e. with high values of dN/dS) tended to exhibit higher levels of caste and behavior bias, (Figure 4E,F; Figure S8E,F), with dN/dS estimated between each focal species (*A. mellifera* and *M. pharaonis*) and a congeneric outgroup (*A. cerana* and *M. chinense*, respectively).

We also detected strong negative correlations between gene-wise expression and overall caste or behavior bias, (correlation of expression and ant caste bias: rho = -0.640; expression and honey bee caste bias: rho = -0.676; expression and ant behavior bias: rho = -0.504; expression and honey bee behavior bias: rho = -0.569; P < 0.001 for all; N = 10520 in ants, N = 10011 in honey bees), where we expression was averaged across the relevant comparisons, analogous to our calculation of overall caste or behavior bias. However, our results are not solely driven by expression levels. We performed partial correlation analysis accounting for expression to test for the conditional effect of evolutionary age, connectivity, tissue specificity, and evolutionary rate on caste or behavior bias. The relationship between caste/behavior bias and evolutionary age, evolutionary rate, connectivity, and tissue-specificity generally all remained significant and in the same direction when expression was accounted for (Table S11). The two exceptions were the relationship between ant behavior bias and evolutionary age and honey bee connectivity and
caste bias (Table S11). However, when we removed the abdominal contribution to caste bias and
to expression, the relationship between honey bee connectivity and caste bias remained significant
and negative (Table S11), revealing that the relationship between connectivity and caste bias was
universal except in the case of highly connected abdominally caste-biased genes, as were present
in our gene coexpression analysis (Figure 2B). Together, these results indicate that each gene’s
likelihood of recruitment to regulatory networks underlying caste or worker division of labor is
positively influenced by expression plasticity and position in overall regulatory networks as well
as the evolutionary rate of coding sequences and how recently such genes have arisen across
evolution.

Discussion

In this study we present the most comprehensive transcriptomic dataset to date of caste-
based division of labor for species representing two independent origins of eusociality. We find
generally low overlap between honey bees and ants in the thousands of genes that are
differentially expressed between queen and worker castes (Figure 1A), and between nurses and
foragers within the worker caste (Figure 1B), indicating that the independent evolution of these
plastic phenotypes in honey bees and ants have largely involved different genes. The one notable
exception is in the abdomen, where we find that about a third (~1500/4500) of genes
differentially expressed between queens and workers are shared between honey bees and ants.
We find that genes with conserved queen bias in ants and honey bees also tend to be female-
biased in D. melanogaster, indicating that these conserved caste-biased genes are derived from
ancient plastically expressed genes underlying sexual dimorphism (Figure 3). Outside of this
conserved core set of ancient genes associated with female reproduction, the majority of genes
associated with caste and worker division of labor showed plastic expression in multiple contexts
(e.g., between developmental stages, tissues, and castes) and were relatively young, rapidly
evolving, and loosely connected in regulatory networks.

In the long-extinct solitary ancestors of social insects, female behaviors such as egg-laying and brood care are thought to have been mechanistically linked within an ovarian cycle such that brood care followed egg-laying (West-Eberhard 1987; Amdam et al. 2006). In extant eusocial societies, these traits are hypothesized to have been de-coupled such that distinct individuals (queens and workers, respectively) independently express pathways associated with egg-laying and brood care (West-Eberhard 1987; Amdam et al. 2006). Crucially, this implies that these pathways existed in solitary ancestors prior to the evolution of eusociality, and reproductive division of labor is derived from these pathways. Consistent with this hypothesis, we found that shared queen-biased abdominal expression in the pharaoh ant and the honey bee is associated with female-biased expression in fruit flies (Figure 3D). Furthermore, we identified a module of co-expressed genes specifically associated with queen abdominal expression in pharaoh ants and honey bees, which were characterized by a core of genes with shared queen-upregulation and known reproductive or maternal function, including vitellogenin and its receptor as well as vasa and smaug (Figure 2A,B; Tables S8,S9). While this use of an ancient set of genes underlying sexual dimorphism for social insect caste dimorphism is striking, it is not surprising, nor is it surprising that the dominant signature of these reproductive-associated genes occurs in abdominal tissues: the fundamental difference between the queen and worker caste is reproduction, and reproductive organs are found in the abdomen (note that pharaoh ant workers
completely lack reproductive organs and are obligately sterile, while honey bee workers are
facultatively sterile and have greatly reduced ovaries relative to queens).

While our results point to a general genetic link between caste dimorphism and sexual
dimorphism, a similar link may exist between caste determination and sex determination.
Mechanisms of sex determination acting early in development are hypothesized to have been co-
opted for the regulation of caste (Klein et al. 2016) and division of labor (Johnson & Cameron
Jasper 2016) by evolving sensitivity to stimuli such as nutrition. Because both queens and
workers are female in eusocial Hymenoptera (Wilson 1971), caste is a sex-limited polyphenism.
Other novel sex-limited polyphenisms such as horns in beetles (Kijimoto et al. 2012) and wing
patterning in butterflies (Kunte et al. 2014) as well as facultative reproduction by honey bee
workers (Velasque et al. 2018) are regulated by the sex determination gene doublesex. In this
way, the ancient developmental switch controlling sex may have been co-opted to determine
alternate developmental trajectories of queens and workers, and subsequent cascades of genes
involved in reproductive physiology of females versus males may have been used to drive the
differentiation of females into highly reproductive queens versus sterile workers.

The tasks performed by workers (specifically, nursing versus foraging) represent a plastic
phenotype that changes over the course of the worker’s adult lifetime (Mikheyev & Linksvayer
2015). This behavioral change is known to be accompanied by a wide range of physiological
changes and is regulated at least in part by conserved physiological pathways, for example, those
involving insulin signaling, juvenile hormone, and vitellogenin (Gospocic et al. 2017; Smith et
al. 2008; Robinson 1987). However, we identified few genes that were commonly differentially
expressed between nurses and foragers in honey bees and pharaoh ants (Figure 1B), and the
The proportion of shared genes was much lower in comparison to genes underlying abdominal differences between queens and workers. Nonetheless, we identified a number of GO terms associated with metabolism as well as development in each species (Tables S6, S7), which is consistent with the notion that the transition from nurse to forager is essentially a developmental process, and that common molecular pathways may provide the raw genetic material for social evolution (Berens et al. 2015; Kapheim et al. 2015).

While conserved factors or pathways appear to play important roles in aspects of caste development and function as well as the transition from nursing to foraging, our results and other studies indicate that the majority of the full transcriptomic architecture associated with caste and age polyethism is not shared between species (Mikheyev & Linksvayer 2015; Ferreira et al. 2013; Berens et al. 2015; Morandin et al. 2016). Part of this lineage-specific architecture is derived from expression differences in conserved genes, as about half the differentially expressed genes we identified had an ortholog in both species but did not exhibit shared expression bias (Figure 1A,B), and the degree of caste-biased expression in particular tissues or stages was generally only weakly correlated among orthologs of honey bees and ants (Figure S1). However, for about 40% of differentially expressed genes, one-to-one orthology could not be determined due to apparent duplication or the complete lack of an ortholog, reflecting the importance of taxonomically-restricted genes for the evolution of eusociality (Sumner 2014; Simola et al. 2013; Jasper et al. 2015; Johnson & Tsutsui 2011; Patalano et al. 2015). Additionally, we found that orthologs with shared abdominal worker bias between honey bees and pharaoh ants were on average evolutionarily younger (i.e. more taxonomically-restricted) than those with shared queen bias (Figure 1C), in line with previous studies (Warner et al. 2017;...
Based on our research and others', it is likely that layered on top of a minority of shared ancient genes associated with sexual dimorphism, the bulk of the genetic architecture underlying caste-based division labor is made up of plastically expressed genes with non-conserved sequence or expression (Ferreira et al. 2013; Berens et al. 2015; Kapheim et al. 2015).

Our results indicate that expression plasticity, both in ancient and taxonomically-restricted genes, is a key feature of the genetic architecture underlying caste and worker division of labor. Overall caste bias across 1-1 orthologs was correlated between *A. mellifera* and *M. pharaonis*, where overall caste bias represents the likelihood of a gene to exhibit expression bias in any comparison between queens and workers (Figure S6A). This indicates that genes that showed high plasticity for caste-biased expression in one species also showed high plasticity for caste-biased expression in the other species. Similarly, overall behavior bias (overall bias between nurses and foragers) was correlated between species (Figure S6B), and caste and behavior bias were themselves correlated (Figure S7). These results indicate that genes with plastic expression in multiple contexts (across tissues, development, behavior, and caste) are consistently recruited for caste and behavior in both honey bees and ants, consistent with the notion that pre-existing plasticity may be utilized for novel polyphenisms (Moczek et al. 2011; Hunt et al. 2011; Leichty et al. 2012).

Consistent with theoretical predictions and empirical results from other studies, we also found that genes with high overall caste and/or behavior bias tended to be loosely connected, plastically expressed, evolutionarily young, and rapidly evolving (Figure 4, Figure S8). Genes underlying conditionally (i.e. plastically) expressed traits are expected to experience relaxed
selection and evolve rapidly at the sequence level (Van Dyken & Wade 2010; Snell-Rood et al. 2010; Kawecki et al. 1997), a prediction that has been empirically supported in a number of organisms (Demuth & Wade 2007; Cruickshank & Wade 2008; Schrader et al. 2017; Purandare et al. 2014; Snell-Rood et al. 2011; Pespeni et al. 2017; Hunt et al. 2011; Leichty et al. 2012).

Similarly, tissue specificity and connectivity affect evolutionary rate because highly pleiotropic (highly connected, not tissue-specific) genes experience enhanced selective constraint (Liao et al. 2006; Mank & Ellegren 2009) (Hahn & Kern 2005; Fraser et al. 2002). Based on these relationships, it is likely that expression plasticity associated with plastic traits (e.g., caste and worker age polyethism) may initially arise largely neutrally as a result of relaxed selective constraint (Morandin et al. 2017; Schrader et al. 2017).

The characteristics we have associated with caste or behavior bias (expression plasticity, evolutionary rate, evolutionary age) form a suite of characteristics which may not be possible to tease apart. For example, across organisms genes identified as taxonomically-restricted tend to evolve more rapidly than older genes (Albà & Castresana 2005; Cai & Petrov 2010; Wolf et al. 2009; Cai et al. 2006), but rapidly evolving genes can mistakenly be taken for taxonomically-restricted genes by BLASTP due to the rapid divergence of their amino acid sequence (Moyers & Zhang 2016; Moyers & Zhang 2015; Moyers & Zhang 2018). However, there doesn’t appear to be a readily available method to distinguish between weak conservation of coding sequence (rapid evolution) and a lack of conservation of the entire sequence itself (taxonomically-restricted genes) (Moyers & Zhang 2018). Similarly, as discussed above, plasticity in expression profile is also clearly linked to evolutionary rate. As such, it may be most useful to think of these characteristics as components of a broad class of gene. Evolutionary developmental biology
emphasizes the action of a certain type of gene for the evolution of morphological novelty: core
to networks, highly pleiotropic, and slowly evolving. However, the emerging picture is that the
evolution of other traits such as social innovation may largely utilize genes on the opposite end
of the spectrum: genes that are peripheral elements to networks and exhibit high levels of
divergence in expression and sequence between species (Mikheyev & Linksvayer 2015; Jasper et
al. 2015).

Sociogenomic studies often seek to find common molecular mechanisms underlying
social evolution in diverse lineages (Robinson et al. 2005; Toth & Robinson 2007). Indeed, we
found approximately 1,500 genes associated with abdominal caste differences in both honey bees
and pharaoh ants, and these genes largely appear to be associated with queen reproduction.
However, it is important to note that other than the striking phenomenon of strong queen-worker
dimorphism that characterizes the reproductive caste system in both honey bees and ants, the
details of social living are remarkably different between honey bees and ants. For example,
honey bee workers fly to forage while ant workers forage on foot (Wilson 1971). This is
reflected in the large number of genes differentially expressed in the thorax between honey bee
foragers and largely flightless nurses compared to few differences between ant nurses and
foragers (Figure 1B), both of which lack wings. Brood care behavior is also quite different:
honey bees rear brood in cells in which they deposit food, while ants pile brood together and feed
larvae via mouth-to-mouth liquid exchange (trophallaxis) or by manually inserting solid food
into larval mouths (Wilson 1971). In light of these and other phenotypic differences, it probably
shouldn’t be surprising that we largely see distinct sets of genes associated with division of labor
in honey bees and ants, consistent with previous research considering the genetic basis of social
behavior across different social insect lineages (Berens et al. 2015; Kapheim et al. 2015; Ferreira et al. 2013; Woodard et al. 2011; Woodard et al. 2014).

Conclusions

Our study suggests that the bulk of the genetic architecture underlying social insect caste-based division of labor varies between lineages. While some sets of conserved pathways such as the insulin/TOR pathways are certainly important across all social insect lineages, it also seems clear that a large suite of additional genes must be necessary for the full articulation of dramatic polyphenisms associated with complex societies (Trible & Kronauer 2017). It is likely that a relatively small number of core conserved genes exist as upstream hubs in regulatory networks, and layered on top of this core are a myriad of taxonomically-restricted genes as well as conserved genes with lineage-specific expression patterns (Mikheyev & Linksvayer 2015; Johnson 2018; Jasper et al. 2015). This is consistent with models for the evolution of hierarchical developmental gene regulatory networks, whereby a small number of genes that act upstream to initiate gene cascades (e.g., to set up body-patterning) are very highly conserved, while batteries of later-acting genes that are relatively downstream in regulatory networks, are much more evolutionarily labile, and are largely responsible for lineage-specific features (Erwin & Davidson 2009; Davidson & Peter 2015). Recent studies have made progress elucidating the function of several core candidate genes and gene pathways that regulate caste and the division of labor by manipulating individual genes (Chandra et al. 2018; Simola et al. 2016; Gospocic et al. 2017; Mutti et al. 2011). While individual core genes are no doubt important, association and linkage mapping studies have shown that most phenotypes (e.g., disease traits) are very highly
polygenic, such that even though only a small number of genes can be identified with large effect sizes and relatively straightforward causative molecular paths, many more genes act peripherally, with smaller effects and through more circuitous routes to profoundly influence trait expression (Boyle et al. 2017). Therefore, a complementary goal of candidate gene-based research is to use unbiased approaches to identify the full suite of genes underlying phenotypic innovations of interest in multiple independent lineages and subsequently to ask what proportion of genes show conserved function across lineages. Large-scale transcriptomic studies such as ours are an invaluable starting point to achieve this goal. However, transcriptomic studies can only hint at the relative functional importance of genes, and future studies will have to assess the functional importance of various components of the transcriptomic architecture, including the relative importance of conserved versus lineage-specific genetic mechanisms for social innovation, and phenotypic innovation more broadly.

Methods

Study Design

We collected parallel time series RNA-seq data of caste development in the pharaoh ant Monomorium pharaonis and the honey bee Apis mellifera, including seven developmental stages (egg, five larval stages, one pupal stage) plus each of three adult body segments in each species (see Table S1 for list of all samples). We separated adults into the three main body segments (head, mesosoma, and metosoma) upon sample collection and sequenced pools of each part separately. For convenience, we refer to these segments as “head”, “thorax”, and “abdominal” tissues throughout. We sequenced whole embryos and whole bodies of larvae and pupae. Each
sample represents a pool of ten individuals taken from the same colony, and replicate samples of
the same type represent pools of individuals taken from different colonies.

RNA extraction, sequencing, aligning to genomes

We isolated RNA using Trizol reagents. We performed cDNA synthesis and library preparation
were performed using the protocol described by Aird et al. (Aird et al. 2017), except that the
input RNA was 50ng and the cycle number of cDNA amplification increased to sixteen. To
compare sample quality across the experiment and test our ability to detect lowly-expressed
genes, we added ERCC92 (Thermo Fisher Scientific Inc.) spike-in mixes to total RNA prior to
amplification. We pooled libraries with an equal amount of cDNA and sequenced single-end for
50 cycles in Illumina Hiseq 2500. We aligned reads to reference genomes using Bowtie2
(Langmead & Salzberg 2012). *A. mellifera* reads were aligned to NCBI gene models, genome
version 4.5, and *M. pharaonis* reads were aligned to NCBI gene models, version 2.0. We
estimated read count and transcripts per million (TPM) using RSEM (Li & Dewey 2011).

Identification of orthologs

To identify orthologs between *A. mellifera* and *M. pharaonis*, we started with a curated
orthology map of *Aculeata* species from OrthoDB9 (Zdobnov et al 2016). We downloaded
amino acid sequences for each species from RefSeq (O’Leary et al. 2016). We associated
transcripts with OrthoDB9 protein names using BLASTp (E-value $10^{-10}$) and identified the
*Aculeata* ortholog group matched by each gene based on the identified BLASTp hits. In this
way, we identified one-to-one, one-to-many, and many-to-many orthologous groups between *A.
mellifera* and *M. pharaonis*. For direct comparison of the species, we restricted our analysis to
one-to-one orthologs (i.e. genes for which only one gene from each species matches the given OrthoDB9 ortholog group). We identified three-way 1-1-1 orthologs between *A. mellifera*, *M. pharaonis*, and *Drosophila melanogaster* using a similar procedure based on *Endopterygota* orthology groups from OrthoDB9.

**Differential Expression Analysis**

To identify genes associated with caste development, we performed differential expression analysis between queens and workers at each developmental stage and tissue, separately for each species. We constructed GLM-like models including replicate and caste and identified genes associated with caste (FDR < 0.1) at each stage using EdgeR (Robinson et al. 2010). To identify caste-associated genes in adults, we treated each stage separately and compared mated queens with nurses and foragers. Similarly, we identified sex-associated genes by comparing males and mated queens.

**Gene coexpression analysis**

In contrast to many network methods which assess gene-gene relationships across all samples, biclustering seeks to identify a group of genes which are coexpressed (i.e. exhibit concerted expression changes) across a subset of sample types (Tanay et al. 2002). Given that our data contained a large number of sample types, we reasoned that we could employ biclustering to identify groups of genes particularly associated with a given sample type. We performed plaid clustering, one of the top performing biclustering algorithms in a recent survey (Oghabian et al. 2014). We used the R package biclust to implement clustering (Kaiser & Leisch 2008). Plaid
clustering models expression level for each gene as a function of bicluster weights, where only biclusters containing the gene contribute to predicted expression level (Turner et al. 2005; Lazzeroni & Owen 2002). The algorithm iteratively constructs layers containing samples and genes and retains layers that improve the model fit, where layers represent biclusters.

Plaid clustering is non-deterministic and individual biclusters are not found in every iteration of clustering. To define a reasonable ensemble of biclusters, we performed clustering 1000 times separately for each species, using inverse hyperbolic sine transformed tpm (transcripts per million) (Brawand et al. 2011). While a large number of interesting bicluster definitions are possible, we decided to identify biclusters that consistently contained all queen abdomen samples to focus our investigation on the tissue that exhibited the strongest signature of caste bias. Specifically, we extracted biclusters containing all three mature queen abdomen samples and no more than three other samples total (note that honey bee queen abdomen samples clustered with egg samples, while pharaoh ant queen samples did not cluster with egg samples).

Because the same genes were not always present in such a bicluster, we tabulated the number of queen abdomen biclusters each gene was found in and retained genes present in a higher proportion of biclusters than a given cut-off, determined by inspection of frequency distributions of bicluster presence. In pharaoh ants, we found a large set of genes present in greater than 90% of queen abdomen biclusters, and we retained these genes for further analysis (N = 1039 genes; Figure S10A, i.e. the same set of genes was repeatedly found). In contrast, honey bee queen abdomen biclusters tended to contain one of two groups of genes, as the frequency of presence in the bicluster peaks at 60% and 30% (Figure S10B). Out of 1245 genes present in greater than 60% of the identified biclusters, 877 were differentially expressed and
upregulated in queen abdomens relative to worker (also note that this set of genes exhibited much higher expression in eggs than the latter set). In contrast, out of 1057 genes present in 25-35% of biclusters, 611 out of were differentially expressed and upregulated in worker abdomens, compared to 47 upregulated in queen abdomens. Therefore, it is clear that the more common bicluster represents genes associated with queen abdomens, so we retained this set of genes for further analysis (N = 1245 genes).

We proceeded with our analysis using these identified sets of genes, which we term modules associated with queen abdominal expression. We calculated connectivity in the module (i.e. intra-module connectivity) as the sum of pairwise Pearson correlations, where correlation values are raised to the sixth power, the standard value for unsigned weighted gene coexpression networks (Langfelder & Horvath 2008). A major goal of gene coexpression analysis is the identification of hub genes, genes central to networks that are strongly associated to relevant traits (Zhang & Horvath 2005). To this end, we conservatively identified hub genes associated with queen abdominal expression as genes with intra-module connectivity in at least the 90th percentile and abdominal log\textsubscript{2} fold-change values greater than 2 (representing a 4-fold increase in expression in queen relative to worker abdomens).

Phylostratigraphy

We estimated the evolutionary age of each gene using phylostratigraphy. Phylostratigraphy groups genes into hierarchical age categories based on identifiable orthology (using BLASTp) (Domazet-Lošo & Tautz 2010; Domazet-Lošo et al. 2007). For example, genes found in ants and honey bees but not in non-aculeate hymenopterans would be labeled “Aculeata” genes, while
genes shared between vertebrates and insects would be labeled “Bilateria”. For our purposes, we decided to focus on the difference between “ancient” genes, which we defined as displaying orthology with non-insect animals, and a number of hierarchical younger categories: “insect”, “Hymenoptera”, “Aculeata”, “ant”, “bee”, and “novel” (where “ant” refers to genes found in *M. pharaonis* and other ants but not in any other species, “bee” refers to genes found in *A. mellifera* and other bees but not in other species, and “novel” refers to a gene found only in *A. mellifera* or *M. pharaonis*).

A key component of phylostratigraphy is the creation of a BLAST database in which to identify orthologs (Domazet-Lošo & Tautz 2010; Domazet-Lošo et al. 2007). Because we largely planned to focus on younger age categories, we constructed a protein database containing all annotated hymenopteran genomes (48 total). We added to this group ten non-hymenopteran insect genomes and ten non-arthropod genomes (see Table S12 for a full list of included genomes). Therefore, a gene labeled as “ancient” displayed a significant BLASTp hit to one of the ten non-arthropod genomes. While phylostratigraphy typically employs an extremely large database containing all available representative taxa, we reasoned that for our study resolution between categories such as “Bilateria” and “Eukaryota” was unnecessary. Furthermore, adding extraneous genomes effectively dilutes the database, such that more similarity is needed to pass an E-value threshold. Because we included only a sample of non-hymenopteran genomes, we were therefore able to stringently identify orthologs (E-value $10^{-10}$ in comparison to a typical value of $10^{-5}$ (Drost et al. 2015)) and accurately place them along the hymenoptera phylogeny.

Comparison to *Drosophila melanogaster*
We reasoned that conserved patterns of sex-bias we witnessed between ants and honey bees were likely highly conserved throughout deep evolutionary time. To address this hypothesis, we used available MODENCODE RNA-seq data on male and female whole bodies of *D. melanogaster* (*Gerstein et al. 2014*). Male and female samples came from 5- and 30-day old flies (one replicate for each sex). While much recent work has detailed tissue-specific (and even cell-specific) expression in *D. melanogaster*, we reasoned that whole bodies (lacking the same body parts as we used in this study) would give us reasonable overall levels of sex bias to compare our data to.

We aligned reads to the current release of the *D. melanogaster* genome available on NCBI (*assembly release 6 plus ISO*) using the same pipeline as described for *A. mellifera* and *M. pharaonis*, and we performed differential expression analysis between males and females as described above.

**Calculation of overall bias**

Previous work estimated the overall degree to which genes exhibited morph bias in RNA-seq data among four morphs by calculating the Euclidean distance of log$_2$ fold-change between each comparison (Schrader et al. 2017). Following this reasoning, we calculated overall caste bias as the Euclidean distance of log$_2$ fold-change between queens and workers at each larval developmental stage (L2-L5), pupae, and each adult tissue. Similarly, we calculated overall behavior bias as the Euclidean distance of log$_2$ fold-change between nurses and foragers in each tissue. To perform partial correlations controlling for magnitude of gene expression with overall bias as the dependent variable, we calculated expression analogously, as the Euclidean distance of log$_{10}$ counts-per-million at each stage/tissue tested.
Estimation of tissue specificity

While comparing expression between castes across the entirety of development in many tissues was unfeasible for this study, a previous study performed RNA-sequencing on twelve tissues in *A. mellifera* in worker nurses and foragers (Jasper et al. 2015). We downloaded the available data and aligned reads to the current version of the *A. mellifera* genome, using the pipeline described above. To classify genes by their tissue specificity, we calculated $\tau$, a commonly used metric of expression specificity (Yanai et al. 2005). $\tau$ ranges from 0 to 1, where 0 indicates genes are ubiquitously expressed and 1 indicates genes are exclusively expressed in one tissue.

Estimation of $dN/dS$

We estimated evolutionary rate using $dN/dS$, the ratio of non-synonymous to synonymous nucleotide changes. We estimated pairwise $dN/dS$ between each focal species and a second closely related species. For *A. mellifera*, we compared protein-coding sequences to *A. cerana*, and for *M. pharaonis*, we compared protein-coding sequences to *Solenopsis invicta*. We identified 1-1 orthologs using OrthoDB9, as described above. For each 1-1 ortholog pair, we selected the longest transcript associated with the gene for each pair of species. We aligned orthologous protein sequences using ClustalW (Larkin et al. 2007), derived nucleotide alignments from protein alignments using pal2nal (Suyama et al. 2006), and estimated pairwise $dN/dS$ of nucleotide alignments using PAML, package codeml (Yang 2007).

Gene Ontology Analysis
We performed Gene Set Enrichment Analysis (GSEA) using the R package topGO (Alexa & Rahnenfuhrer 2010). Rather than relying on automatically annotated Gene Ontology (GO) terms, we utilized the well-curated *D. melanogaster* gene ontology database, downloaded from FlyBase (Gramates et al. 2016). We performed GSEA analysis on genes with 1-1 orthologs identified previously, associating the *D. melanogaster* GO terms to *A. mellifera* and *M. pharaonis* orthologs. We identified GO terms associated caste-biased or behavior-biased genes using the P-value of differential expression between queens and workers or nurses and foragers. We identified GO terms associated with overall caste or behavior bias using the value of overall caste bias or behavior bias as input. We identified enriched terms with P-value < 0.05.

**General statistical analyses**

We performed all statistical analyses and constructed plots in R, version 3.4.0 (R Core Team 2017), aided by the packages “plyr” (Wickham & Others 2011), “ggplot2” (Wickham 2016), “gridExtra”, version 2.3 (Auguie 2012) and “cowplot”, version 0.9.3 (Wilke 2016). To test for the effect of phylostrata on caste or behavior bias, we constructed a generalized linear model using the Gamma distribution and treating phylostrata as an ordinal variable. To test for the effects of connectivity and evolutionary rate (dN/dS) on caste or behavior bias while controlling for expression, we employed partial correlation analysis using the package “ppcor” (Kim 2015).

**Data availability**
All data and scripts required to generate figures, tables, and perform statistical analyses are available on Github: https://github.com/oist/devnetwork. Raw reads are deposited at DNA Data Bank of Japan.

Acknowledgements

We thank Luigi Pontieri for photographs of pharaoh ants, and Alex Wild for photographs of honey bees. We thank Chao Tong for compiling Hymenopteran genomes for phylostratigraphy analysis and for comments on the manuscript. We thank Junhyong Kim, Mia Levine, Justin Walsh, and Rohini Singh for helpful comments and discussion. This work was funded by the National Science Foundation (grant number IOS-1452520 to TAL) United States Department of Agriculture National Institute of Food and Agriculture (grant number 2014-67013-21725 to TAL), and subsidy funding from Okinawa Institute of Technology to ASM. MRW was funded by The National Science Foundation also funded MRW (DGE-1321851).

References

Albà, M.M. & Castresana, J., 2005. Inverse relationship between evolutionary rate and age of mammalian genes. Molecular biology and evolution, 22(3), pp.598–606. Available at: http://dx.doi.org/10.1093/molbev/msi045.

Alexa, A. & Rahnenfuhrer, J., 2010. topGO: enrichment analysis for gene ontology. R package version, 2(0). Available at: http://bioconductor.uib.no/2.7/bioc/html/topGO.html.

Amdam, G.V. et al., 2006. Complex social behaviour derived from maternal reproductive traits. Nature, 439(7072), pp.76–78. Available at: http://dx.doi.org/10.1038/nature04340.

Amdam, G.V. et al., 2004. Reproductive ground plan may mediate colony-level selection effects on individual foraging behavior in honey bees. Proceedings of the National Academy of Sciences of the United States of America, 101(31), pp.11350–11355. Available at: http://dx.doi.org/10.1073/pnas.0403073101.
Auguie, B., 2012. gridExtra: functions in Grid graphics. *R package version 0.9*, 1.

Babonis, L.S., Martindale, M.Q. & Ryan, J.F., 2016. Do novel genes drive morphological novelty? An investigation of the nematosomes in the sea anemone *Nematostella vectensis*. *BMC evolutionary biology*, 16(1), p.114. Available at: http://dx.doi.org/10.1186/s12862-016-0683-3.

Barchuk, A.R., Bitondi, M.M.G. & Simões, Z.L.P., 2002. Effects of juvenile hormone and ecdysone on the timing of vitellogenin appearance in hemolymph of queen and worker pupae of *Apis mellifera*. *Journal of insect science*, 2, p.1.

Beldade, P. & Brakefield, P.M., 2002. The genetics and evo-devo of butterfly wing patterns. *Nature reviews. Genetics*, 3(6), pp.442–452. Available at: http://dx.doi.org/10.1038/nrg818.

Benoit, B. et al., 2009. An essential role for the RNA-binding protein Smaug during the *Drosophila* maternal-to-zygotic transition. *Development*, 136(6), pp.923–932. Available at: http://dx.doi.org/10.1242/dev.031815.

Berens, A.J., Hunt, J.H. & Toth, A.L., 2015. Comparative transcriptomics of convergent evolution: different genes but conserved pathways underlie caste phenotypes across lineages of eusocial insects. *Molecular biology and evolution*, 32(3), pp.690–703. Available at: http://dx.doi.org/10.1093/molbev/msu330.

Boyle, E.A., Li, Y.I. & Pritchard, J.K., 2017. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell*, 169(7), pp.1177–1186. Available at: http://dx.doi.org/10.1016/j.cell.2017.05.038.

Branstetter, M.G. et al., 2017. Phylogenomic Insights into the Evolution of Stinging Wasps and the Origins of Ants and Bees. *Current biology: CB*, 27(7), pp.1019–1025. Available at: http://dx.doi.org/10.1016/j.cub.2017.03.027.

Brawand, D. et al., 2011. The evolution of gene expression levels in mammalian organs. *Nature*, 478(7369), pp.343–348. Available at: http://dx.doi.org/10.1038/nature10532.

Cai, J.J. et al., 2006. Accelerated evolutionary rate may be responsible for the emergence of lineage-specific genes in ascomycota. *Journal of molecular evolution*, 63(1), pp.1–11. Available at: http://dx.doi.org/10.1007/s00239-004-0372-5.

Cai, J.J. & Petrov, D.A., 2010. Relaxed purifying selection and possibly high rate of adaptation in primate lineage-specific genes. *Genome biology and evolution*, 2, pp.393–409. Available at: http://dx.doi.org/10.1093/gbe/evq019.

Carroll, S.B., 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell*, 134(1), pp.25–36. Available at: http://dx.doi.org/10.1016/j.cell.2008.06.030.
Chandra, V. et al., 2018. Social regulation of insulin signaling and the evolution of eusociality in ants. *Science*, 361(6400), pp.398–402. Available at: http://dx.doi.org/10.1126/science.aar5723.

Chen, S., Krinsky, B.H. & Long, M., 2013. New genes as drivers of phenotypic evolution. *Nature reviews. Genetics*, 14(9), pp.645–660. Available at: http://dx.doi.org/10.1038/nrg3521.

Cruickshank, T. & Wade, M.J., 2008. Microevolutionary support for a developmental hourglass: gene expression patterns shape sequence variation and divergence in *Drosophila*. *Evolution & development*, 10(5), pp.583–590. Available at: http://doi.wiley.com/10.1111/j.1525-142X.2008.00273.x.

Davidson, E.H., 2010. *The Regulatory Genome: Gene Regulatory Networks In Development And Evolution*. Academic Press, San Diego, California.

Davidson, E.H. & Peter, I.S., 2015. *Genomic control process*, Academic Press, San Diego, California.

Demuth, J.P. & Wade, M.J., 2007. Maternal expression increases the rate of bicoid evolution by relaxing selective constraint. *Genetica*, 129(1), pp.37–43. Available at: http://dx.doi.org/10.1007/s10709-006-0031-4.

Domazet-Lošo, T., Brajković, J. & Tautz, D., 2007. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends in genetics: TIG*, 23(11), pp.533–539. Available at: http://dx.doi.org/10.1016/j.tig.2007.08.014.

Domazet-Lošo, T. & Tautz, D., 2010. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature*, 468(7325), pp.815–818. Available at: http://dx.doi.org/10.1038/nature09632.

Drost, H.-G. et al., 2015. Evidence for active maintenance of phylotranscriptomic hourglass patterns in animal and plant embryogenesis. *Molecular biology and evolution*, 32(5), pp.1221–1231. Available at: http://dx.doi.org/10.1093/molbev/msv012.

Erwin, D.H. & Davidson, E.H., 2009. The evolution of hierarchical gene regulatory networks. *Nature reviews. Genetics*, 10(2), pp.141–148. Available at: http://dx.doi.org/10.1038/nrg2499.

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. *Molecular ecology*, 23(1), pp.151–161. Available at: http://onlinelibrary.wiley.com/doi/10.1111/mec.12490/full.

Ferreira, P.G. et al., 2013. Transcriptome analyses of primitively eusocial wasps reveal novel
insights into the evolution of sociality and the origin of alternative phenotypes. *Genome biology*, 14(2), p.R20. Available at: http://dx.doi.org/10.1186/gb-2013-14-2-r20.

Fraser, H.B. et al., 2002. Evolutionary rate in the protein interaction network. *Science*, 296(5568), pp.750–752. Available at: http://dx.doi.org/10.1126/science.1068696.

Gerstein, M.B. et al., 2014. Comparative analysis of the transcriptome across distant species. *Nature*, 512(7515), pp.445–448. Available at: http://dx.doi.org/10.1038/nature13424.

Gospocic, J. et al., 2017. The Neuropeptide Corazonin Controls Social Behavior and Caste Identity in Ants. *Cell*, 170(4), pp.748–759.e12. Available at: http://dx.doi.org/10.1016/j.cell.2017.07.014.

Gramates, L.S. et al., 2016. FlyBase at 25: looking to the future. *Nucleic acids research*, 45(D1), pp.D663-D671. Available at: https://doi.org/10.1093/nar/gkw1016.

Hahn, M.W. & Kern, A.D., 2005. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. *Molecular biology and evolution*, 22(4), pp.803–806. Available at: http://dx.doi.org/10.1093/molbev/msi072.

Hilgers, L. et al., 2018. Novel Genes, Ancient Genes, and Gene Co-Option Contributed to the Genetic Basis of the Radula, a Molluscan Innovation. *Molecular biology and evolution*, 35(7), pp.1638–1652. Available at: http://dx.doi.org/10.1093/molbev/msy052.

Hoekstra, H.E. & Coyne, J.A., 2007. The locus of evolution: evo devo and the genetics of adaptation. *Evolution; international journal of organic evolution*, 61(5), pp.995–1016. Available at: http://dx.doi.org/10.1111/j.1558-5646.2007.00105.x.

Hölldobler, B. & Wilson, E.O., 2009. *The Superorganism: The Beauty, Elegance, and Strangeness of Insect Societies*, W. W. Norton & Company, New York City, New York.

Hunt, B.G. et al., 2011. Relaxed selection is a precursor to the evolution of phenotypic plasticity. *Proceedings of the National Academy of Sciences of the United States of America*, 108(38), pp.15936–15941. Available at: http://dx.doi.org/10.1073/pnas.1104825108.

Hwang, J.S. et al., 2007. The evolutionary emergence of cell type-specific genes inferred from...
the gene expression analysis of Hydra. *Proceedings of the National Academy of Sciences of the United States of America*, 104(37), pp.14735–14740. Available at: http://dx.doi.org/10.1073/pnas.0703331104.

Jasper, W.C. et al., 2015. Large-scale coding sequence change underlies the evolution of postdevelopmental novelty in honey bees. *Molecular biology and evolution*, 32(2), pp.334–346. Available at: http://dx.doi.org/10.1093/molbev/msu292.

Johnson, B.R., 2018. Taxonomically Restricted Genes Are Fundamental to Biology and Evolution. *Frontiers in genetics*, 9, p.407. Available at: https://www.frontiersin.org/article/10.3389/fgene.2018.00407.

Johnson, B.R. & Cameron Jasper, W., 2016. Complex patterns of differential expression in candidate master regulatory genes for social behavior in honey bees. *Behavioral ecology and sociobiology*, 70(7), pp.1033–1043. Available at: https://doi.org/10.1007/s00265-016-2071-9

Johnson, B.R. & Linksvayer, T.A., 2010. Deconstructing the superorganism: social physiology, groundplans, and sociogenomics. *The Quarterly review of biology*, 85(1), pp.57–79. Available at: https://doi.org/10.1086/650290.

Johnson, B.R. & Tsutsui, N.D., 2011. Taxonomically restricted genes are associated with the evolution of sociality in the honey bee. *BMC genomics*, 12, p.164. Available at: http://dx.doi.org/10.1186/1471-2164-12-164.

Kaiser, S. & Leisch, F., 2008. A toolbox for bicluster analysis in R. Available at: https://doi.org/10.5282/ubm/epub.3293

Kapheim, K.M. et al., 2015. Social evolution. Genomic signatures of evolutionary transitions from solitary to group living. *Science*, 348(6239), pp.1139–1143. Available at: http://dx.doi.org/10.1126/science.aaa4788.

Kawecki, T.J., Barton, N.H. & Fry, J.D., 1997. Mutational collapse of fitness in marginal habitats and the evolution of ecological specialisation. *Journal of evolutionary biology*, 10(3), pp.407–429. Available at: https://doi.org/10.1046/j.1420-9101.1997.10030407.x.

Khalturin, K. et al., 2008. A novel gene family controls species-specific morphological traits in Hydra. *PLoS biology*, 6(11), p.e278. Available at: http://dx.doi.org/10.1371/journal.pbio.0060278.

Khalturin, K. et al., 2009. More than just orphans: are taxonomically-restricted genes important in evolution? *Trends in genetics: TIG*, 25(9), pp.404–413. Available at: http://dx.doi.org/10.1016/j.tig.2009.07.006.

Khila, A. & Abouheif, E., 2010. Evaluating the role of reproductive constraints in ant social
evolution. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 365(1540), pp.617–630. Available at: http://dx.doi.org/10.1098/rstb.2009.0257.

Kijimoto, T., Moczek, A.P. & Andrews, J., 2012. Diversification of doublesex function underlies morph-, sex-, and species-specific development of beetle horns. *Proceedings of the National Academy of Sciences of the United States of America*, 109(50), pp.20526–20531. Available at: http://dx.doi.org/10.1073/pnas.1118589109.

Kim, S., 2015. `ppcor`: An R Package for a Fast Calculation to Semi-partial Correlation Coefficients. *Communications for statistical applications and methods*, 22(6), pp.665–674. Available at: http://dx.doi.org/10.5351/CSAM.2015.22.6.665.

Klein, A. et al., 2016. Evolution of Social Insect Polyphenism Facilitated by the Sex Differentiation Cascade. *PLoS genetics*, 12(3), p.e1005952. Available at: http://dx.doi.org/10.1371/journal.pgen.1005952.

Kunte, K. et al., 2014. doublesex is a mimicry supergene. *Nature*, 507(7491), pp.229–232. Available at: http://dx.doi.org/10.1038/nature13112.

Langfelder, P. & Horvath, S., 2008. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics*, 9, p.559. Available at: http://dx.doi.org/10.1186/1471-2105-9-559.

Langmead, B. & Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9(4), pp.357–359. Available at: http://dx.doi.org/10.1038/nmeth.1923.

Larkin, M.A. et al., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21), pp.2947–2948. Available at: https://doi.org/10.1093/bioinformatics/btm404

Lazzeroni, L. & Owen, A., 2002. Plaid models for gene expression data. *Statistica Sinica*, 12(1), pp.61–86. Available at: http://www.jstor.org/stable/24307036.

Leichty, A.R. et al., 2012. Relaxed genetic constraint is ancestral to the evolution of phenotypic plasticity. *Integrative and comparative biology*, 52(1), pp.16–30. Available at: http://dx.doi.org/10.1093/icb/ics049.

Liao, B.-Y., Scott, N.M. & Zhang, J., 2006. Impacts of gene essentiality, expression pattern, and gene compactness on the evolutionary rate of mammalian proteins. *Molecular biology and evolution*, 23(11), pp.2072–2080. Available at: http://dx.doi.org/10.1093/molbev/msl076.

Libbrecht, R. et al., 2013. Interplay between insulin signaling, juvenile hormone, and vitellogenin regulates maternal effects on polyphenism in ants. *Proceedings of the National Academy of Sciences of the United States of America*, 110(27), pp.11050–11055. Available at: http://dx.doi.org/10.1073/pnas.1221781110.
Li, B. & Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with
or without a reference genome. *BMC bioinformatics*, 12, p.323. Available at:
http://dx.doi.org/10.1186/1471-2105-12-323.

Mank, J.E. & Ellegren, H., 2009. Are sex-biased genes more dispensable? *Biology letters*, 5(3),
p.409–412. Available at: http://dx.doi.org/10.1098/rsbl.2008.0732.

Mével-Ninio, M. et al., 1995. ovo, a Drosophila gene required for ovarian development, is
specifically expressed in the germline and shares most of its coding sequences with
shavenbaby, a gene involved in embryo patterning. *Mechanisms of development*, 49(1),
p.83–95. Available at:
http://www.sciencedirect.com/science/article/pii/0925477394003057.

Mikheyev, A.S. & Linksvayer, T.A., 2015. Genes associated with ant social behavior show
distinct transcriptional and evolutionary patterns. *eLife*, 4, e04775. Available at:
http://dx.doi.org/10.7554/eLife.04775.

Milde, S. et al., 2009. Characterization of taxonomically restricted genes in a phylum-restricted
cell type. *Genome biology*, 10(1), p.R8. Available at: http://dx.doi.org/10.1186/gb-2009-10-1-r8.

Moczek, A.P. et al., 2011. The role of developmental plasticity in evolutionary innovation.
*Proceedings. Biological sciences / The Royal Society*, 278(1719), pp.2705–2713. Available
at: http://dx.doi.org/10.1098/rspb.2011.0971.

Morandin, C. et al., 2016. Comparative transcriptomics reveals the conserved building blocks
involved in parallel evolution of diverse phenotypic traits in ants. *Genome biology*, 17, p.43.
Available at: http://dx.doi.org/10.1186/s13059-016-0902-7.

Morandin, C., Mikheyev, A.S. & Pedersen, J.S., 2017. Evolutionary constraints shape caste-
specific gene expression across 15 ant species. Available at:
http://onlinelibrary.wiley.com/doi/10.1111/evo.13220/full.

Moyers, B.A. & Zhang, J., 2016. Evaluating Phylostratigraphic Evidence for Widespread De
Novo Gene Birth in Genome Evolution. *Molecular biology and evolution*, 33(5), pp.1245–
1256. Available at: http://dx.doi.org/10.1093/molbev/msw008.

Moyers, B.A. & Zhang, J., 2015. Phylostratigraphic bias creates spurious patterns of genome
evolution. *Molecular biology and evolution*, 32(1), pp.258–267. Available at:
http://dx.doi.org/10.1093/molbev/msu286.

Moyers, B.A. & Zhang, J., 2018. Toward Reducing Phylostratigraphic Errors and Biases.
*Genome biology and evolution*, 10(8), pp.2037–2048. Available at:
http://dx.doi.org/10.1093/gbe/evy161.
Müller, G.B., 2007. Evo–devo: extending the evolutionary synthesis. *Nature reviews. Genetics*, 8, p.943. Available at: http://dx.doi.org/10.1038/nrg2219.

Muller, G.B. & Wagner, G.P., 1991. Novelty in Evolution: Restructuring the Concept. *Annual review of ecology and systematics*, 22(1), pp.229–256. Available at: https://doi.org/10.1146/annurev.es.22.110191.001305.

Mutti, N.S. et al., 2011. IRS and TOR nutrient-signaling pathways act via juvenile hormone to influence honey bee caste fate. *The Journal of experimental biology*, 214(Pt 23), pp.3977–3984. Available at: http://dx.doi.org/10.1242/jeb.061499.

Nielsen, C. & Martinez, P., 2003. Patterns of gene expression: homology or homocracy? *Development genes and evolution*, 213(3), pp.149–154. Available at: http://dx.doi.org/10.1007/s00427-003-0301-4.

O’Connell, L.A. & Hofmann, H.A., 2012. Evolution of a vertebrate social decision-making network. *Science*, 336(6085), pp.1154–1157. Available at: http://dx.doi.org/10.1126/science.1218889.

Oghabian, A. et al., 2014. Biclustering methods: biological relevance and application in gene expression analysis. *PloS one*, 9(3), p.e90801. Available at: http://dx.doi.org/10.1371/journal.pone.0090801.

Okada, Y. et al., 2017. Social dominance alters nutrition-related gene expression immediately: transcriptomic evidence from a monomorphic queenless ant. *Molecular ecology*, 26(11), pp.2922–2938. Available at: http://dx.doi.org/10.1111/mec.13989.

O’Leary, N.A. et al., 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic acids research*, 44(D1), pp.D733–45. Available at: http://dx.doi.org/10.1093/nar/gkv1189.

Page, R.E., Jr & Amdam, G.V., 2007. The making of a social insect: developmental architectures of social design. *BioEssays: news and reviews in molecular, cellular and developmental biology*, 29(4), pp.334–343. Available at: http://dx.doi.org/10.1002/bies.20549.

Patalano, S. et al., 2015. Molecular signatures of plastic phenotypes in two eusocial insect species with simple societies. *Proceedings of the National Academy of Sciences of the United States of America*, 112(45), pp.13970–13975. Available at: http://dx.doi.org/10.1073/pnas.1515937112.

Peacock, A.D. & Baxter, A.T., 1950. Studies in Pharaoh’s ant, *Monomorium pharaonis* (L.), 3: life history. *Entomologist’s Monthly Magazine*, 86, pp.171–178.

Pespeni, M.H., Ladner, J.T. & Moczek, A.P., 2017. Signals of selection in conditionally expressed genes in the diversification of three horned beetle species. *Journal of...*
Prud’homme, B. et al., 2006. Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. *Nature*, 440(7087), pp.1050–1053. Available at: http://dx.doi.org/10.1038/nature04597.

Purandare, S.R. et al., 2014. Accelerated evolution of morph-biased genes in pea aphids. *Molecular biology and evolution*, 31(8), pp.2073–2083. Available at: http://dx.doi.org/10.1093/molbev/msu149.

Rittschof, C.C. & Robinson, G.E., 2016. Behavioral Genetic Toolkits: Toward the Evolutionary Origins of Complex Phenotypes. *Current topics in developmental biology*, 119, pp.157–204. Available at: http://dx.doi.org/10.1016/bs.ctdb.2016.04.001.

Robinson, G.E., 1987. Regulation of honey bee age polyethism by juvenile hormone. *Behavioral ecology and sociobiology*, 20(5), pp.329–338. Available at: https://link.springer.com/article/10.1007/BF00300679 [Accessed August 26, 2017].

Robinson, G.E., Grozinger, C.M. & Whitfield, C.W., 2005. Sociogenomics: social life in molecular terms. *Nature reviews. Genetics*, 6(4), pp.257–270. Available at: http://dx.doi.org/10.1038/nrg1575.

Robinson, M.D., McCarthy, D.J. & Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), pp.139–140. Available at: http://dx.doi.org/10.1093/bioinformatics/btp616.

Santos, M.E. et al., 2017. Taxon-restricted genes at the origin of a novel trait allowing access to a new environment. *Science*, 358(6361), pp.386–390. Available at: http://dx.doi.org/10.1126/science.aan2748.

Saul, M.C. et al., 2018. Cross-species systems analysis of evolutionary toolkits of neurogenomic response to social challenge. *Genes, brain, and behavior*, p.e12502. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/gbb.12502.

Schmidt, A.M. et al., 2011. Queen–worker caste ratio depends on colony size in the pharaoh ant (*Monomorium pharaonis*). *Insectes sociaux*, 58(2), pp.139–144. Available at: https://link.springer.com/article/10.1007/s00040-010-0126-x [Accessed April 26, 2017].

Schrader, L., Helanterä, H. & Oettler, J., 2017. Accelerated evolution of developmentally biased genes in the tetraphenic ant *Cardiocondyla obscurior*. *Molecular biology and evolution*, 34(3), pp.535–544. Available at: http://dx.doi.org/10.1093/molbev/msw240.

Shubin, N., Tabin, C. & Carroll, S., 2009. Deep homology and the origins of evolutionary novelty. *Nature*, 457(7231), pp.818–823. Available at:
Simola, D.F. et al., 2016. Epigenetic (re)programming of caste-specific behavior in the ant *Camponotus floridanus*. *Science*, 351(6268), p.aac6633. Available at: http://dx.doi.org/10.1126/science.aac6633.

Simola, D.F. et al., 2013. Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. *Genome research*, 23(8), pp.1235–1247. Available at: http://dx.doi.org/10.1101/gr.155408.113.

Smith, C.R. et al., 2008. Genetic and genomic analyses of the division of labour in insect societies. *Nature reviews. Genetics*, 9(10), pp.735–748. Available at: http://dx.doi.org/10.1038/nrg2429.

Snell-Rood, E.C. et al., 2011. Developmental decoupling of alternative phenotypes: insights from the transcriptomes of horn-polyphenic beetles. *Evolution; international journal of organic evolution*, 65(1), pp.231–245. Available at: http://dx.doi.org/10.1111/j.1558-5646.2010.01106.x.

Snell-Rood, E.C. et al., 2010. Toward a population genetic framework of developmental evolution: the costs, limits, and consequences of phenotypic plasticity. *BioEssays: news and reviews in molecular, cellular and developmental biology*, 32(1), pp.71–81. Available at: http://dx.doi.org/10.1002/bies.200900132.

Sullivan, J.P. et al., 2000. Juvenile Hormone paces behavioral development in the adult worker honey bee. *Hormones and behavior*, 37(1), pp.1–14. Available at: http://www.sciencedirect.com/science/article/pii/S0018506X99915520.

Sumner, S., 2014. The importance of genomic novelty in social evolution. *Molecular ecology*, 23(1), pp.26–28. Available at: http://dx.doi.org/10.1111/mec.12580.

Suyama, M., Torrents, D. & Bork, P., 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic acids research*, 34(Web Server issue), pp.W609–12. Available at: http://dx.doi.org/10.1093/nar/gkl315.

Tanay, A., Sharan, R. & Shamir, R., 2002. Discovering statistically significant biclusters in gene expression data. *Bioinformatics*, 18 Suppl 1, pp.S136–44. Available at: https://www.ncbi.nlm.nih.gov/pubmed/12169541.

Tautz, D. & Domazet-Lošo, T., 2011. The evolutionary origin of orphan genes. *Nature reviews. Genetics*, 12(10), pp.692–702. Available at: http://dx.doi.org/10.1038/nrg3053.

Toth, A.L. et al., 2010. Brain transcriptomic analysis in paper wasps identifies genes associated with behaviour across social insect lineages. *Proceedings. Biological sciences / The Royal Society*, 277(1691), pp.2139–2148. Available at: http://dx.doi.org/10.1098/rspb.2010.0090.
Toth, A.L. & Robinson, G.E., 2007. Evo-devo and the evolution of social behavior. *Trends in genetics: TIG*, 23(7), pp.334–341. Available at: http://dx.doi.org/10.1016/j.tig.2007.05.001.

Trible, W. & Kronauer, D.J.C., 2017. Caste development and evolution in ants: it’s all about size. *The Journal of experimental biology*, 220(Pt 1), pp.53–62. Available at: http://dx.doi.org/10.1242/jeb.145292.

Turner, H., Bailey, T. & Krzanowski, W., 2005. Improved biclustering of microarray data demonstrated through systematic performance tests. *Computational statistics & data analysis*, 48(2), pp.235–254. Available at: http://www.sciencedirect.com/science/article/pii/S0167947304000295.

Van Dyken, J.D. & Wade, M.J., 2010. The genetic signature of conditional expression. *Genetics*, 184(2), pp.557–570. Available at: http://dx.doi.org/10.1534/genetics.109.110163.

Velasque, M., Qiu, L. & Mikheyev, A.S., 2018. The Doublesex sex determination pathway regulates reproductive division of labor in honey bees. bioRxiv, p.314492. Available at: https://doi.org/10.1101/314492.

Wagner, G.P., 2014. *Homology, Genes, and Evolutionary Innovation*, Princeton University Press, Princeton, NJ.

Warner, M.R., Kovaka, K. & Linksvayer, T.A., 2016. Late-instar ant worker larvae play a prominent role in colony-level caste regulation. *Insectes sociaux*, 63(4), pp.575–583. Available at: https://doi.org/10.1007/s00040-016-0501-3

Warner, M.R., Mikheyev, A.S. & Linksvayer, T.A., 2017. Genomic signature of kin Selection in an ant with obligately sterile workers. *Molecular biology and evolution*, 34(7), pp.1780–1787. Available at: http://dx.doi.org/10.1093/molbev/msx123.

West-Eberhard, M.J., 1987. "Flexible strategy and social evolution" in *Animal societies. Theories and facts*, edited by Ito, Yoshiaki, Brown, Janine L., and Kikkawa, J., 35–51. Japan Scientific Societies Press, Tokyo, Japan.

Wheeler, D.E., 1986. Developmental and physiological determinants of caste in social Hymenoptera: evolutionary implications. *The American naturalist*, 128(1), pp.13–34. Available at: http://www.jstor.org/stable/2461282.

Wheeler, D.E., Buck, N. & Evans, J.D., 2006. Expression of insulin pathway genes during the period of caste determination in the honey bee, Apis mellifera. *Insect molecular biology*, 15(5), pp.597–602. Available at: http://dx.doi.org/10.1111/j.1365-2583.2006.00681.x.

Wickham, H., 2016. *ggplot2: Elegant Graphics for Data Analysis*, Springer, Houston, Texas.

Wickham, H. & Others, 2011. The split-apply-combine strategy for data analysis. *Journal of
statistical software, 40(1), pp.1–29. Available at:
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.182.5667&rep=rep1&type=pdf.

Wilke, C.O., 2016. cowplot: Streamlined Plot Theme and Plot Annotations for “ggplot2”. R package version 0.7.0.

Wilson, E.O., 1971. The insect societies. The insect societies, Harvard University Press, Boston, Massachusetts.

Wilson, G.A. et al., 2007. Large-scale comparative genomic ranking of taxonomically restricted genes (TRGs) in bacterial and archaeal genomes. PloS one, 2(3), p.e324. Available at:
http://dx.doi.org/10.1371/journal.pone.0000324.

Wilson, G.A. et al., 2005. Orphans as taxonomically restricted and ecologically important genes. Microbiology, 151(Pt 8), pp.2499–2501. Available at:
http://dx.doi.org/10.1099/mic.0.28146-0.Wittkopp, P.J. & Kalay, G., 2011. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nature reviews. Genetics, 13(1), pp.59–69. Available at: http://dx.doi.org/10.1038/nrg3095.

Wolf, Y.I. et al., 2009. The universal distribution of evolutionary rates of genes and distinct characteristics of eukaryotic genes of different apparent ages. Proceedings of the National Academy of Sciences of the United States of America, 106(18), pp.7273–7280. Available at:
http://dx.doi.org/10.1073/pnas.0901808106.

Woodard, S.H. et al., 2011. Genes involved in convergent evolution of eusociality in bees. Proceedings of the National Academy of Sciences of the United States of America, 108(18), pp.7472–7477. Available at: http://dx.doi.org/10.1073/pnas.1103457108.

Woodard, S.H. et al., 2014. Molecular heterochrony and the evolution of sociality in bumblebees (Bombus terrestris). Proceedings. Biological sciences / The Royal Society, 281(1780), p.20132419. Available at: http://dx.doi.org/10.1098/rspb.2013.2419.

Wray, G.A., 2007. The evolutionary significance of cis-regulatory mutations. Nature reviews. Genetics, 8(3), pp.206–216. Available at: http://dx.doi.org/10.1038/nrg2063.

Wray, G.A. et al., 2003. The evolution of transcriptional regulation in eukaryotes. Molecular biology and evolution, 20(9), pp.1377–1419. Available at:
http://dx.doi.org/10.1093/molbev/msg140.

Yanai, I. et al., 2005. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics , 21(5), pp.650–659. Available at: http://dx.doi.org/10.1093/bioinformatics/bti042.

Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Molecular biology and evolution, 24(8), pp.1586–1591. Available at: http://dx.doi.org/10.1093/molbev/msm088.
Zhang, B. & Horvath, S., 2005. A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology*, 4, p.17. Available at: http://dx.doi.org/10.2202/1544-6115.1128.
Figure 1. Number of differentially expressed genes (FDR < 0.1) between A) queens and workers and B) nurses and foragers at each developmental stage or tissue in *M. pharaonis* (left) and *A. mellifera* (right). “Head”, “thorax”, and “abdomen” refer to body segments of adults, while pupa and larva refers to whole bodies. Genes are divided by color into categories based on gene-gene orthology between the two species. Insets show the proportion of each category of gene out of all differentially expressed genes at that stage. C) Genes with conserved queen bias (upregulated in queen abdomens of both species) are comprised of proportionally ancient genes than those with conserved worker bias or genes with non-conserved bias *: the category “larva” represents differential expression across larvae of all stages for which caste can be identified (second to fifth larval stage).
Figure 2. Abdominal caste bias (log₂ fold change queen versus worker) is correlated with connectivity within the queen-abdomen specific bicluster (i.e. module) in A) ants (rho = 0.536, P < 0.001) and B) honey bees (rho = 0.617, P < 0.001), indicating that centrally located genes in the bicluster are highly caste-biased. Genes upregulated in queens are in red, while genes upregulated in workers are in blue. Connectivity is proportional to the most highly connected gene in the module. Furthermore, connectivity within the queen abdominal module is higher in genes found in the module for both species (shared) versus genes found in the module for only one species (not shared) in C) ants and D) honey bees. ***P < 0.001 (Wilcoxon test)
Figure 3. Caste bias is derived from sex bias. Abdominal caste bias (queen vs worker log$_2$ fold-change) is correlated to abdominal sex bias (queen vs male log$_2$ fold-change) in A) *M. pharaonis* (rho = 0.715, P < 0.001) and B) *A. mellifera* (rho = 0.774, P < 0.001) and abdominal sex bias is correlated between the two species (rho = 0.280, P < 0.001) (C). Genes are colored by conservation of differential expression in abdomens, with red indicating genes upregulated in queens in both species, blue indicating genes upregulated in workers of both species, and grey indicating genes that exhibited non-conserved expression patterns. Finally, genes with conserved queen bias tend to be female-biased in *D. melanogaster* based on whole-body adult samples while genes with conserved worker bias tend to be male-biased in *D. melanogaster* (likely reflecting down-regulation in females).
Figure 4. Genes that exhibit more caste bias across body segments and developmental stages have younger estimated evolutionary ages (A,B) and tend to be loosely connected (C,D; ant: ρ = -0.159, P < 0.001; honey bee: ρ = -0.090, P < 0.001) and rapidly evolving (E,F; ant: ρ = 0.157, P < 0.001; honey bee: ρ = 0.240, P < 0.001). “Overall caste bias” combines queen/worker log₂ fold-change values across all development stages and adult body segments. Connectivity is calculated using all samples and genes and scaled proportionally to the highest value.
Figure S1. Pearson correlation of log-fold change between queens and workers as measured at each stage or tissue in *M. pharaonis* and *A. mellifera* for each 1-1 ortholog (N = 7640). Error bars indicate pearson correlation 95% confidence intervals. In (A), the category “larva*” represents differential expression across larval stages, while in (B) each larval stage (L2-L5) is plotted individually.
Figure S2. Pearson correlation of log-fold change between nurses and foragers as measured in each tissue in *M. pharaonis* and *A. mellifera* for each 1-1 ortholog (N = 7640). Error bars indicate pearson correlation 95% confidence intervals.
Figure S3. Log$_2$ fold-change at each stage/tissue in each phyllostrata category. Positive values indicating higher expression in queens compared to workers. Log$_2$ fold-change has been adjusted relative to the median value at that stage/tissue, in order to compare across tests. “Ancient” genes indicate any genes shared beyond insects (i.e. with vertebrates).
Figure S4. Pearson correlation of caste (queen/worker) and sex (queen/male) expression bias in ants and honey bees. Error bars represent Pearson correlation 95% confidence intervals. Correlations are significant in all cases (P < 0.001), but abdominal correlations are strongest.
**Figure S5.** Number of times each gene is upregulated in queen and workers across all comparisons (larva, pupa, and adult head, thorax, and abdomen). Color brightness is logarithmically proportional to the number of genes in each cell. N = 10804 genes (*M. pharaonis*); 11775 genes (*A. mellifera*).
Figure S6. Overall caste bias (A) and overall behavior bias (B) is correlated between ants and honey bees. “Overall” bias refers to the euclidean distance of all log2 fold-change values (queens/worker for caste, nurses/foragers for behavior). Red line is trendline of linear model; Spearman correlation P < 0.001 in all cases.
Figure S7. Overall caste bias and overall behavior bias were correlated within A) ants and B) honey bees. “Overall” bias refers to the euclidean distance of all log2 fold-change values (queens/worker for caste, nurses/foragers for behavior). Red line is trendline of linear model; Spearman correlation P < 0.001 in all cases.
Figure S8. Genes that exhibit more behavior bias across body segments have younger estimated evolutionary ages (A,B) and tend to be loosely connected (C,D; ant: rho = -0.099, P < 0.001; honey bee: rho = -0.157, P < 0.001) and rapidly evolving (E,F; ant: rho = 0.079, P < 0.001; honey bee: rho = 0.226, P < 0.001). “Overall behavior bias” combines nurse forager log2 fold-change values across all adult body segments. Connectivity is calculated using all samples and genes and scaled proportionally to the highest value.
Figure S9. Genes exhibiting more behavior bias tend to be tissue-specific. There was a positive correlation (Spearman correlation, \(P < 0.001\) in each case) between caste/behavior bias and tissue specificity, where tissue specificity (\(\tau\)) is estimated using data from 12 honey bee tissues. \(\tau = 1\) indicates a gene is expressed in only one tissue, while lower values indicate genes are more ubiquitously (i.e. evenly) expressed across tissues.
Figure S10. Histogram of the frequency with which genes were placed in the queen abdomen bicluster (out of 1000 runs). Plaid biclustering is a non-deterministic process, so different sets of genes can be present in each run. In ants (A), 1039 genes were present in >90% of queen abdomen biclusters and retained for further analysis. There are two peaks in the frequency distribution for honey bees (B). The lower frequency peak is made up of worker-associated genes (downregulated in queen abdomens) while the higher frequency peak (~60%) is made up of queen-associated genes. We retained genes with >60% frequency for further analysis.
Table S1. Full listing of sample types and number of each sample collected. “L1” and “L2” refers to larvae of the first and second stage, etc. We began caste-specific sampling at stage two because caste is determined and regulated in *M. pharaonis* by the end of the first larval instar (Warner et al. 2016). After the first larval instar in *M. pharaonis*, worker-destined larvae can be distinguished from reproductive-destined larvae, which include male-destined and queen-destined larvae (Peacock & Baxter 1950). As such, our “queen-destined” ant larvae samples likely contain some male-destined larvae, but the proportion is expected to be low, as the sex ratio is known to be heavily queen-biased (Schmidt et al. 2011). Sex and caste are both known in *A. mellifera* larvae, as individuals are reared in separate cells (Wilson 1971).

| Species | Stage | Tissue | Caste | Type | Replicates |
|---------|-------|--------|-------|------|------------|
| ant     | egg   | whole body | N/A   | N/A  | 6          |
| ant     | L1    | whole body | N/A   | N/A  | 6          |
| ant     | L2    | whole body | queen | N/A  | 3          |
| ant     | L2    | whole body | worker | N/A | 3          |
| ant     | L3    | whole body | queen | N/A  | 3          |
| ant     | L3    | whole body | worker | N/A | 3          |
| ant     | L4    | whole body | queen | N/A  | 3          |
| ant     | L4    | whole body | worker | N/A | 3          |
| ant     | L5    | whole body | queen | N/A  | 3          |
| ant     | L5    | whole body | worker | N/A | 3          |
| ant     | pupa  | whole body | queen | N/A  | 3          |
| ant     | pupa  | whole body | worker | N/A | 3          |
| ant     | pupa  | whole body | male   | N/A  | 3          |
| ant     | adult | head     | queen | virgin | 3         |
| ant     | adult | thorax   | queen | virgin | 3         |
| ant     | adult | abdomen  | queen | virgin | 3         |
| ant     | adult | head     | queen | mated | 3         |
| ant     | adult | thorax   | queen | mated | 3         |
| ant     | adult | abdomen  | queen | mated | 3         |
| ant     | adult | head     | worker | nurse | 3         |
| ant     | adult | thorax   | worker | nurse | 3         |
| ant     | adult | abdomen  | worker | nurse | 3         |
| ant     | adult | head     | worker | forager | 3         |
| ant     | adult | thorax   | worker | forager | 3         |
| ant     | adult | abdomen  | worker | forager | 3         |
| ant     | adult | head     | male   | N/A  | 3         |
| ant     | adult | thorax   | male   | N/A  | 3         |
| ant     | adult | abdomen  | male   | N/A  | 3         |
| honey bee | egg   | whole body | N/A   | N/A  | 3          |
| honey bee | L1    | whole body | N/A   | N/A  | 3          |
| honey bee | L2    | whole body | queen | N/A  | 3          |
| honey bee | L2    | whole body | worker | N/A | 3          |
| honey bee | L3    | whole body | queen | N/A  | 3          |
| honey bee | L3    | whole body | worker | N/A | 3          |
| honey bee | L4    | whole body | queen | N/A  | 3          |
| honey bee | L4    | whole body | worker | N/A | 3          |
| honey bee | L5    | whole body | queen | N/A  | 3          |
| honey bee | L5    | whole body | worker | N/A | 4          |
| honey bee | pupa  | whole body | queen | N/A  | 5          |
| honey bee | pupa  | whole body | worker | N/A | 3          |
| honey bee | pupa  | whole body | male   | N/A  | 3          |
| honey bee | adult | head     | queen | virgin | 3         |
| honey bee | adult | thorax   | queen | virgin | 3         |
| honey bee | adult | abdomen  | queen | virgin | 3         |
| honey bee | adult | thorax   | queen | virgin | 3         |
| honey bee | adult | head     | queen | mated | 3         |
| honey bee | adult | thorax   | queen | mated | 3         |
| honey bee | adult | abdomen  | queen | mated | 3         |
| honey bee | adult | head     | worker | nurse | 3         |
| honey bee | adult | thorax   | worker | nurse | 3         |
| honey bee | adult | abdomen  | worker | nurse | 3         |
| honey bee | adult | head     | worker | forager | 3         |
| honey bee | adult | thorax   | worker | forager | 3         |
| honey bee | adult | abdomen  | worker | forager | 3         |
| honey bee | adult | head     | male   | N/A  | 3         |
| honey bee | adult | thorax   | male   | N/A  | 3         |
| honey bee | adult | abdomen  | male   | N/A  | 3         |
| stage/tissue | species    | total DEGs | queen associated | worker associated |
|-------------|------------|------------|------------------|-------------------|
| L2          | ant        | 119        | 31               | 88                |
| L3          | ant        | 818        | 490              | 328               |
| L4          | ant        | 81         | 65               | 16                |
| L5          | ant        | 757        | 437              | 320               |
| pupa        | ant        | 290        | 88               | 202               |
| head        | ant        | 741        | 420              | 321               |
| thorax      | ant        | 1327       | 695              | 632               |
| abdomen     | ant        | 4395       | 2711             | 1684              |
| larva_overall | ant     | 361        | 241              | 120               |
| L2          | honey bee  | 136        | 60               | 76                |
| L3          | honey bee  | 117        | 28               | 89                |
| L4          | honey bee  | 724        | 224              | 500               |
| L5          | honey bee  | 1009       | 540              | 469               |
| pupa        | honey bee  | 245        | 163              | 82                |
| head        | honey bee  | 1144       | 717              | 427               |
| thorax      | honey bee  | 1369       | 721              | 648               |
| abdomen     | honey bee  | 5352       | 2769             | 2583              |
| larva_overall | honey bee | 473        | 176              | 297               |

**Table S2.** Number of differentially expressed genes (DEGs) between queens and workers for each comparison (FDR < 0.1). “L2”, “L3”, etc refer to the 2nd and 3rd larval stage, respectively, while “larva_overall” is the result of differential expression with caste as main effect across all larval samples. Differentially expressed genes are divided into “queen associated”, which exhibited higher expression in queens, and “worker associated”, which exhibited higher expression in workers. Differential expression analysis performed with N = 10804 (ant) and 11775 (honey bee) genes.
| GO ID   | Term                                      | P       | Stage/Tissue |
|---------|-------------------------------------------|---------|--------------|
| GO:0007265 | Ras protein signal transduction            | 0.00026 | larva        |
| GO:0046578 | regulation of Ras protein signal transduction | 0.00029 | larva        |
| GO:0051056 | regulation of small GTPase mediated signal transduction | 0.00048 | larva        |
| GO:0007284 | small GTPase mediated signal transduction  | 0.00071 | larva        |
| GO:0060628 | regulation of ER to Golgi vesicle-mediated transport | 0.00071 | larva        |
| GO:0006970 | response to osmotic stress                 | 0.00089 | pupa         |
| GO:0009651 | response to salt stress                    | 0.00287 | pupa         |
| GO:0019432 | triglyceride biosynthetic process           | 0.00287 | pupa         |
| GO:0046460 | neutral lipid biosynthetic process         | 0.00287 | pupa         |
| GO:0046463 | acylglycerol biosynthetic process           | 0.00287 | pupa         |
| GO:0072525 | pyridine-containing compound biosynthetic process | 0.0018 | head         |
| GO:0001704 | formation of primary germ layer            | 0.0042  | head         |
| GO:0002098 | RNA wobble uridine modification            | 0.0050  | head         |
| GO:0043086 | negative regulation of catalytic activity  | 0.0071  | head         |
| GO:0010508 | positive regulation of autophagy           | 0.0078  | head         |
| GO:0019362 | pyridine nucleotide metabolic process      | 0.00021 | thorax       |
| GO:0046496 | nicotinamide nucleotide metabolic process  | 0.00021 | thorax       |
| GO:0006733 | oxidoreduction coenzyme metabolic process   | 0.00043 | thorax       |
| GO:0072524 | pyridine-containing compound metabolic process | 0.00054 | thorax       |
| GO:0006739 | NADP metabolic process                     | 0.00088 | thorax       |
| GO:0042445 | hormone metabolic process                  | 7.5e-05 | abdomen      |
| GO:0010817 | regulation of hormone levels               | 0.00041 | abdomen      |
| GO:0042181 | ketone biosynthetic process                | 0.00089 | abdomen      |
| GO:0042180 | cellular ketone metabolic process          | 0.00094 | abdomen      |
| GO:0034754 | cellular hormone metabolic process         | 0.00108 | abdomen      |

**Table S3.** Enriched gene ontology terms based on Gene Set Enrichment Analysis (GSEA) of differential expression between queens and workers in ants. P-value derived from Kolmogorov-Smirnov tests.
### Table S4.

Enriched gene ontology terms based on Gene Set Enrichment Analysis (GSEA) of differential expression between queens and workers in honey bees. P-value derived from Kolmogorov-Smirnov tests.

| GO.ID       | Term                                                | P       | stage/tissue |
|-------------|-----------------------------------------------------|---------|--------------|
| GO:0019722  | calcium-mediated signaling                          | 0.0010  | larva        |
| GO:0046113  | nucleobase catabolic process                        | 0.0021  | larva        |
| GO:0007411  | axon guidance                                       | 0.0024  | larva        |
| GO:0061564  | axon development                                    | 0.0027  | larva        |
| GO:0035039  | male pronucleus assembly                            | 0.0032  | larva        |
| GO:0008544  | epidermis development                               | 0.0053  | pupa         |
| GO:0008286  | insulin receptor signaling pathway                  | 0.0065  | pupa         |
| GO:0034599  | cellular response to oxidative stress               | 0.0097  | pupa         |
| GO:0032869  | cellular response to insulin stimulus               | 0.0106  | pupa         |
| GO:0071375  | cellular response to peptide hormone stimulus       | 0.0106  | pupa         |
| GO:0008610  | lipid biosynthetic process                          | 0.0011  | head         |
| GO:0016070  | RNA metabolic process                               | 0.0018  | head         |
| GO:0030534  | adult behavior                                      | 0.0028  | head         |
| GO:0008344  | adult locomotory behavior                           | 0.0031  | head         |
| GO:0007478  | leg disc morphogenesis                              | 0.0041  | head         |
| GO:0048747  | muscle fiber development                            | 0.011   | thorax       |
| GO:0090254  | cell elongation involved in imaginal disc-derived wing morphogenesis | 0.027 | thorax       |
| GO:0006457  | protein folding                                     | 0.045   | thorax       |
| GO:0071897  | DNA biosynthetic process                            | 0.055   | thorax       |
| GO:0034063  | stress granule assembly                             | 0.071   | thorax       |
| GO:0045887  | positive regulation of synaptic growth at neuromuscular junction | 0.0037 | abdomen      |
| GO:1904398  | positive regulation of neuromuscular junction development | 0.0037 | abdomen      |
| GO:0051965  | positive regulation of synapse assembly             | 0.0116  | abdomen      |
| GO:0030490  | maturation of SSU-RNA                               | 0.0126  | abdomen      |
| GO:0007436  | larval salivary gland morphogenesis                 | 0.0205  | abdomen      |
Table S5. Number of differentially expressed genes (DEGs) between nurses and foragers for each comparison (FDR < 0.1). Differentially expressed genes are divided into “nurse associated”, which exhibited higher expression in nurses, and “forager associated”, which exhibited higher expression in foragers. Differential expression analysis performed with N = 10804 (ant) and 11775 (honey bee) genes.

| stage/tissue | species       | total DEGs | nurse associated | forager associated |
|--------------|---------------|------------|------------------|--------------------|
| head         | ant           | 405        | 314              | 91                 |
| thorax       | ant           | 490        | 305              | 185                |
| abdomen      | ant           | 544        | 341              | 203                |
| head         | honey bee     | 927        | 404              | 523                |
| thorax       | honey bee     | 2519       | 1243             | 1276               |
| abdomen      | honey bee     | 2017       | 1007             | 1010               |
Table S6. Enriched gene ontology terms based on Gene Set Enrichment Analysis (GSEA) of differential expression between nurses and foragers in ants. P-value derived from Kolmogorov-Smirnov tests.

| GO.ID      | Term                                           | P     | tissue |
|------------|------------------------------------------------|-------|--------|
| GO:0048284 | organelle fusion                                | 0.00021 | head  |
| GO:0006629 | lipid metabolic process                          | 0.00138 | head  |
| GO:0048580 | regulation of post-embryonic development         | 0.00186 | head  |
| GO:0044255 | cellular lipid metabolic process                 | 0.00217 | head  |
| GO:0044801 | single-organism membrane fusion                  | 0.00226 | head  |
| GO:0032502 | developmental process                            | 0.00019 | thorax |
| GO:0007525 | somatic muscle development                       | 0.00035 | thorax |
| GO:0044767 | single-organism developmental process            | 0.00041 | thorax |
| GO:0007275 | multicellular organism development               | 0.00041 | thorax |
| GO:0090175 | regulation of establishment of planar polarity   | 0.00043 | thorax |
| GO:0000289 | nuclear-transcribed mRNA poly(A) tail shortening | 0.0017  | abdomen |
| GO:0007006 | mitochondrial membrane organization               | 0.0054  | abdomen |
| GO:0042441 | eye pigment metabolic process                    | 0.0056  | abdomen |
| GO:0043324 | pigment metabolic process involved in development | 0.0056  | abdomen |
| GO:0043474 | pigment metabolic process involved in pigmentation| 0.0056  | abdomen |
| GO ID     | Term                                                        | P   | Stage/Tissue |
|-----------|-------------------------------------------------------------|-----|--------------|
| GO:0019722| calcium-mediated signaling                                  | 0.0010 | larva       |
| GO:0046113| nucleobase catabolic process                                | 0.0021 | larva       |
| GO:0007411| axon guidance                                              | 0.0024 | larva       |
| GO:0061564| axon development                                           | 0.0027 | larva       |
| GO:0035039| male pronucleus assembly                                   | 0.0032 | larva       |
| GO:0008544| epidermis development                                      | 0.00053 | pupa     |
| GO:0008286| insulin receptor signaling pathway                          | 0.00065 | pupa     |
| GO:0034599| cellular response to oxidative stress                      | 0.00097 | pupa     |
| GO:0032869| cellular response to insulin stimulus                       | 0.00106 | pupa     |
| GO:0071375| cellular response to peptide hormone stimulus              | 0.00106 | pupa     |
| GO:0008610| lipid biosynthetic process                                 | 0.00011 | head    |
| GO:0016070| RNA metabolic process                                      | 0.00018 | head    |
| GO:0030534| adult behavior                                             | 0.00028 | head    |
| GO:0008344| adult locomotory behavior                                  | 0.00031 | head    |
| GO:0007478| leg disc morphogenesis                                     | 0.00041 | head    |
| GO:0048747| muscle fiber development                                   | 0.0011 | thorax |
| GO:0090254| cell elongation involved in imaginal disc-derived wing morphogenesis | 0.0027 | thorax |
| GO:0006457| protein folding                                            | 0.0045 | thorax |
| GO:0071897| DNA biosynthetic process                                   | 0.0055 | thorax |
| GO:0034063| stress granule assembly                                    | 0.0071 | thorax |
| GO:0045887| positive regulation of synaptic growth at neuromuscular junction | 0.00037 | abdomen |
| GO:1904398| positive regulation of neuromuscular junction development   | 0.00037 | abdomen |
| GO:0051965| positive regulation of synapse assembly                     | 0.00116 | abdomen |
| GO:0030490| maturation of SSU-RNA                                       | 0.00126 | abdomen |
| GO:0007436| larval salivary gland morphogenesis                         | 0.00205 | abdomen |

**Table S7.** Enriched gene ontology terms based on Gene Set Enrichment Analysis (GSEA) of differential expression between nurses and foragers in honey bees. P-value derived from Kolmogorov-Smirnov tests.
Table S8. Hub genes of the queen abdominal module in ants. Hub genes were defined as genes with intra-modular connectivity in at least the 90th percentile, and \(\log_2\) fold-change (queen/worker) greater than 2.

| Gene      | logFC queen/worker | connectivity | SwissProt                                      |
|-----------|--------------------|--------------|------------------------------------------------|
| LOC105837185 | 9.333              | 0.793        | Leukocyte elastase inhibitor A                 |
| LOC105838268 | 8.610              | 0.790        | Gephyrin                                       |
| LOC105836111 | 8.147              | 0.784        | RCC1 and BTB domain-containing protein 1        |
| LOC105836023 | 7.962              | 0.772        | Histone H2B                                    |
| LOC105834654 | 7.258              | 0.797        | Vitellogenin receptor                           |
| LOC105837528 | 6.871              | 0.848        | Ankyrin-2                                      |
| LOC105838623 | 6.713              | 0.855        | Transcription factor SOX-14                   |
| LOC105829700 | 6.447              | 0.839        | Maternal embryonic leucine zipper kinase       |
| LOC105834441 | 6.307              | 0.864        | Nuclear RNA export factor 1                    |
| LOC105830728 | 6.115              | 0.911        | S-phase kinase-associated protein 2            |
| LOC105837219 | 5.825              | 0.926        | Rac GTPase-activating protein 1                |
| LOC105840991 | 5.777              | 0.973        | Acidic repeat-containing protein               |
| LOC105838786 | 5.389              | 0.824        | Spordin-1                                      |
| LOC105837988 | 5.364              | 0.819        | Multiple PDZ domain protein                    |
| LOC105830806 | 5.311              | 0.833        | Insulin-degrading enzyme                       |
| LOC105836312 | 5.144              | 0.797        | Serine protease nudel                         |
| LOC105836129 | 5.108              | 0.915        | Rho GTPase-activating protein 19               |
| LOC105834586 | 5.078              | 0.906        | Putative bifunctional UDP-N-acetylgallosamine transferase and deubiquitnase ALG13 |
| LOC105832223 | 5.058              | 0.874        | E3 ubiquitin-protein ligase SIAH1             |
| LOC105840292 | 4.974              | 0.785        | Pre-mRNA-splicing factor RBM22                |
| LOC105840093 | 4.833              | 0.897        | ATP-dependent RNA helicase vasa isoform A      |
| LOC105833998 | 4.812              | 0.925        | Piwi-like protein 1                            |
| LOC105839662 | 4.676              | 0.966        | G2/mitotic-specific cyclin-B3                 |
| LOC105828383 | 4.542              | 0.949        | Histone RNA hairpin-binding protein            |
| LOC105830377 | 4.326              | 0.815        | Protein dispatched                            |
| LOC105835848 | 4.292              | 0.807        | Glyoxylate reductase                          |
| LOC105838831 | 4.242              | 0.890        | Broad-complex core protein isoform 6          |
| LOC105832464 | 4.188              | 0.774        | Zinc finger protein 800                       |
| LOC105837226 | 4.096              | 0.883        | DNA repair and recombination protein RAD54-like (Fragment) |
| LOC105834656 | 4.077              | 0.805        | Putative ATP-dependent RNA helicase me31b     |
| Gene       | logFC queen/worker | connectivity | SwissProt                                      |
|------------|--------------------|--------------|------------------------------------------------|
| LOC724752  | 11.662             | 0.746        | E3 ubiquitin-protein ligase TRIM71             |
| LOC410888  | 9.000              | 0.829        | lachesin-like                                  |
| LOC410684  | 7.405              | 0.754        | homeobox protein OTX1 A                       |
| LOC100576333| 6.493             | 0.738        | nucleoradixin-like                            |
| LOC725841  | 6.339              | 0.897        | hyaluronan mediated motility receptor          |
| LOC100577382| 6.272             | 0.896        | targeting protein for Xkp2 homolog            |
| LOC551099  | 6.172              | 0.760        | coiled-coil domain-containing protein 43       |
| LOC724193  | 5.957              | 0.859        | kinesin-like protein KIF18A                    |
| LOC726506  | 5.927              | 0.828        | protein claret segregational                  |
| LOC725920  | 5.829              | 0.875        | vitellogenin receptor                          |
| LOC100576828| 5.826              | 0.843        | protein malstrom 2                             |
| LOC412031  | 5.811              | 0.930        | S-phase kinase-associated protein 2            |
| LOC102666846| 5.705             | 0.942        | cyclin-A2                                      |
| LOC411529  | 5.531              | 0.861        | maternal embryonic leucine zipper kinase-like  |
| LOC410502  | 5.456              | 0.910        | transformation/transcription domain-associated protein |
| LOC410015  | 5.270              | 0.778        | protein LSM14 homolog A                       |
| LOC100578691| 5.186             | 0.909        | rhoGEF domain-containing protein gxclJ-like    |
| LOC100578255| 5.073             | 0.734        | ras GTPase-activating-like protein IQGAP1       |
| LOC552100  | 4.953              | 0.809        | protein ovo                                    |
| LOC100576908| 4.883             | 0.917        | polycomb protein Asx                           |
| LOC551871  | 4.808              | 0.912        | P protein-like                                 |
| LOC411970  | 4.546              | 0.767        | G kinase-anchoring protein 1-like              |
| LOC725606  | 4.498              | 0.758        | serine protease gd                             |
| LOC409092  | 4.486              | 0.850        | enhancer of mRNA-decapping protein 3           |
| LOC409681  | 4.483              | 0.807        | RWD domain-containing protein 1                |
| LOC411809  | 4.460              | 0.780        | enolase-phosphatase E1                         |
| LOC551773  | 4.320              | 0.876        | serine/threonine-protein kinase VRK1-like      |
| LOC413667  | 4.212              | 0.914        | G2/mitotic-specific cyclin-B3                  |
| LOC409472  | 3.995              | 0.975        | protein Smiag homolog 1                       |
| LOC552725  | 3.868              | 0.848        | N-acetylglucosamine-1-phosphotransferase subunits alpha/beta |

**Table S9.** Hub genes of the queen abdominal module in ants. Hub genes were defined as genes with intra-modal connectivity in at least the 90th percentile, and log2 fold-change (queen/worker) greater than 2.
| GO.ID      | Term                                                                 | P       | test          |
|------------|----------------------------------------------------------------------|---------|--------------|
| GO:0010977 | negative regulation of neuron projection development                | 0.00032 | ant caste    |
| GO:0014017 | neuroblast fate commitment                                           | 0.00038 | ant caste    |
| GO:0045165 | cell fate commitment                                                 | 0.00041 | ant caste    |
| GO:0007400 | neuroblast fate determination                                         | 0.00085 | ant caste    |
| GO:0010771 | negative regulation of cell morphogenesis involved in differentiation | 0.00122 | ant caste    |
| GO:0060322 | head development                                                      | 2.3e-07 | ant behavior |
| GO:0007420 | brain development                                                     | 3.6e-07 | ant behavior |
| GO:0016319 | mushroom body development                                             | 1.6e-05 | ant behavior |
| GO:0045165 | cell fate commitment                                                 | 2.1e-05 | ant behavior |
| GO:0007417 | central nervous system development                                    | 4.2e-05 | ant behavior |
| GO:0006650 | glycerophospholipid metabolic process                                 | 0.00046 | bee caste    |
| GO:0051231 | spindle elongation                                                    | 0.00098 | bee caste    |
| GO:0046488 | phosphatidylinositol metabolic process                                | 0.00134 | bee caste    |
| GO:0035050 | embryonic heart tube development                                      | 0.00194 | bee caste    |
| GO:0000022 | mitotic spindle elongation                                            | 0.00222 | bee caste    |
| GO:0035295 | tube development                                                      | 0.00020 | bee behavior |
| GO:0002009 | morphogenesis of an epithelium                                        | 0.00029 | bee behavior |
| GO:0035239 | tube morphogenesis                                                    | 0.00032 | bee behavior |
| GO:0048729 | tissue morphogenesis                                                  | 0.00061 | bee behavior |
| GO:0060562 | epithelial tube morphogenesis                                          | 0.00123 | bee behavior |

**Table S10.** Enriched gene ontology terms based on overall caste or behavior bias in ants and honey bees. GO terms are derived from *D. melanogaster* orthologs. P-value is from gene set enrichment analysis (Kolmogorov–Smirnov test).
| species      | comparison | abdomen included? | variable tested | Spearman rho | P-value   |
|--------------|------------|-------------------|-----------------|--------------|-----------|
| ant          | caste      | yes               | connectivity    | -0.162       | 3.92e-33  |
| ant          | caste      | no                | connectivity    | -0.282       | 1.38e-99  |
| ant          | caste      | yes               | dN/dS           | 0.151        | 6.07e-29  |
| ant          | caste      | no                | dN/dS           | 0.130        | 8.51e-22  |
| ant          | caste      | yes               | evolutionary age| 0.159        | 5.19e-32  |
| ant          | caste      | no                | evolutionary age| 0.161        | 5.96e-33  |
| honey bee    | caste      | yes               | connectivity    | 0.003        | 8.13e-01  |
| honey bee    | caste      | no                | connectivity    | -0.132       | 4.61e-20  |
| honey bee    | caste      | yes               | dN/dS           | 0.155        | 3.15e-27  |
| honey bee    | caste      | no                | dN/dS           | 0.169        | 4.27e-32  |
| honey bee    | caste      | yes               | evolutionary age| 0.166        | 3.72e-31  |
| honey bee    | caste      | no                | evolutionary age| 0.154        | 7.95e-27  |
| honey bee    | caste      | yes               | tau             | 0.409        | 1.33e-193 |
| honey bee    | caste      | no                | tau             | 0.397        | 1.14e-180 |
| ant          | behavior   | yes               | connectivity    | -0.157       | 4.29e-31  |
| ant          | behavior   | no                | connectivity    | -0.119       | 1.24e-18  |
| ant          | behavior   | yes               | dN/dS           | 0.043        | 1.75e-03  |
| ant          | behavior   | no                | dN/dS           | 0.020        | 1.38e-01  |
| ant          | behavior   | yes               | evolutionary age| 0.021        | 1.25e-01  |
| ant          | behavior   | no                | evolutionary age| 0.014        | 3.02e-01  |
| honey bee    | behavior   | yes               | connectivity    | -0.162       | 9.22e-30  |
| honey bee    | behavior   | no                | connectivity    | -0.137       | 1.49e-21  |
| honey bee    | behavior   | yes               | dN/dS           | 0.174        | 7.02e-34  |
| honey bee    | behavior   | no                | dN/dS           | 0.150        | 1.83e-25  |
| honey bee    | behavior   | yes               | evolutionary age| 0.161        | 3.49e-29  |
| honey bee    | behavior   | no                | evolutionary age| 0.126        | 1.95e-18  |
| honey bee    | behavior   | yes               | tau             | 0.328        | 9.55e-121 |
| honey bee    | behavior   | no                | tau             | 0.281        | 5.79e-88  |

Table S11. Partial correlation between connectivity, evolutionary rate (dN/dS), evolutionary age (phylostrata), and tissue-specificity (tau) and caste or behavior bias while accounting for expression. Analysis was performed separately for each species and comparison (i.e. separately for caste bias and behavior bias), as well as while including or excluding abdomen in calculations of caste bias and expression. Connectivity is total connectivity measured across all samples and genes. Phylostrata is a measure of estimated evolutionary age, with higher values indicating younger genes. Tau is the degree to which genes exhibit tissue-specific expression across 12 honey bee tissues (results presented only for honey bees). N = 10520 genes (ants), 10011 genes (honey bees).
**Table S12.** List of species used for phylostratigraphy analysis, with the NCBI Taxonomy ID.

| Species                     | NCBI Taxonomy ID |
|-----------------------------|------------------|
| Acromyrmex echinatior       | 103372           |
| Atta cephalotes             | 12957            |
| Atta colombica              | 520822           |
| Camponotus floridanus       | 104421           |
| Cardiocondyla obscurnor     | 268306           |
| Monomorium pharaonis        | 307668           |
| Ligniphthera humile         | 83485            |
| Lasius niger                | 67767            |
| Harpegnathos saltator       | 610360           |
| Dinoponera quadrispinosa    | 609295           |
| Cyphomyrmex costatus        | 406900           |
| Ocoetaea biroi              | 2018173          |
| Pogononyrmex barbatinus     | 144034           |
| Pseudomyrmex gracilis       | 219809           |
| Solenopsis invicts          | 13686            |
| Trachymyrmex septentrionalis| 34720            |
| Trachymyrmex corneti        | 471704           |
| Trachymyrmex zeteki         | 64791            |
| Vollenhovia emeryi          | 411798           |
| Wasmannia auropunctata      | 64700            |
| Temnothorax curvispinosus   | 300111           |
| Apis cerana                 | 7461             |
| Apis dorsata                | 7462             |
| Apis florea                 | 7463             |
| Apis mellifera              | 7460             |
| Bombus impatiens            | 132113           |
| Bombus terrestris           | 30195            |
| Melipona quadridascataia    | 166423           |
| Eufriesea mexicana          | 516756           |
| Ceratina calcarata          | 156304           |
| Megachile rotundata         | 143995           |
| Habropoda labronica         | 597456           |
| Dufourea novaangiae         | 178035           |
| Polistes canadensis         | 91411            |
| Polistes dominula           | 743375           |
| Ceratocolen solmsi          | 142686           |
| Ceratosolen solmsi marchali | 326594           |
| Copidosoma floridanum       | 29053            |
| Fopius arisanus             | 64638            |
| Microptilus demilbilis      | 69319            |
| Nasonia vitripennis         | 7425             |
| Trichogramma pretiosum      | 7493             |
| Trichomalopsis sarothagae   | 543379           |
| Diachasmia aliceum          | 454923           |
| Oraeus atletirus            | 222816           |
| Althaea rosea               | 37344            |
| Cephus cinclus              | 211228           |
| Necropon lecontei           | 441921           |
| Pediciulus humanus          | 121225           |
| Drosophila melanogaster     | 7227             |
| Aedes aegypti               | 7159             |
| Bombyx mori                 | 7081             |
| Papilio machaon             | 76183            |
| Anopheles gambiase          | 7165             |
| Orthophagus taurus          | 166361           |
| Tribolium castaneum         | 7070             |
| Acyrthosiphon pisum         | 7029             |
| Zootermopsis nevadensis     | 138037           |
| Caenortabotlis elegans      | 6239             |
| Hydro vulgaris              | 6087             |
| Strongylcentronurus purpuratus| 7668         |
| Lottia gigantea             | 225164           |
| Helobella robusta           | 6412             |
| Mus musculus                | 10000            |
| Homo sapiens                | 9006             |
| Xenopus tropicalis          | 8364             |
| Latimeria chalumnae         | 7997             |
| Danio rerio                 | 7985             |