The convergent evolution of eusociality is based on a shared reproductive groundplan plus lineage-specific sets of plastic genes

Michael R. Warner[1], Lijun Qiu[2], Michael J. Holmes[2,3], Alexander S. Mikheyev[2,4], Timothy A. Linksvayer[1]

[1]University of Pennsylvania, Philadelphia, Pennsylvania, USA
[2]Okinawa Institute of Science and Technology, Okinawa, Japan
[3]School of Life and Environmental Science, University of Sydney, Australia
[4]Research 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 remains largely unknown. To elucidate these issues, we characterized caste-specific transcriptomic profiles across development and adult body segments from pharaoh ants (*Monomorium pharaonis*) and honey bees (*Apis mellifera*), representing two independent origins of eusociality. We identified a large shared core of genes upregulated in the abdomens of queen ants and honey bees that also tends to be upregulated in female flies. Outside of this shared core, few genes are differentially expressed in common. Instead, the majority of the thousands of genes underlying the expression of the caste system are plastically-expressed, rapidly evolving, and relatively evolutionary young. Altogether our results show that the convergent evolution of eusociality involves the recruitment of a core reproductive groundplan along with lineage-specific sets of plastic genes.

Significance Statement

The convergent evolution of similar features in different lineages exemplifies how natural selection can produce common adaptive solutions, but whether the same genes are consistently used remains unresolved. Eusociality, defined by caste-based division of labor, convergently evolved in several insect lineages. Previous studies alternately emphasized lineage-specific genes or small sets of shared genes and pathways. We conducted an unprecedented parallel transcriptomic study and identified a large core of ancient reproductive genes that was recruited during the independent evolution of queen and worker castes in ants and honey bees. Outside this shared core, we find thousands more caste-associated genes with distinct features that are not shared, demonstrating that convergent eusocial evolution is based on a combination of shared and lineage-specific genes.

Introduction

The degree to which convergent phenotypic evolution involves the same sets of genes or pathways is a major unanswered question (1). Comparative genomic studies indicate that parallel adaptive changes in the protein-coding sequences of the same genes are frequently associated with the evolution of convergent phenotypes in closely related populations and species (2, 3). Decades of research in evolutionary developmental biology also emphasize that changes in the expression of a relatively small “toolkit” of deeply conserved genes are often associated with convergently evolved phenotypes in distantly related species (4). Alternatively, convergent phenotypic evolution between lineages could involve distinct subsets of genes in each lineage. Taxonomically-restricted genes, genes which are only found in certain lineages (5) have been

shown to be important for lineage-specific evolutionary novelties (6), but their relative contribution to the evolution of convergent phenotypes is unknown.

The evolution of eusociality in several insect lineages (e.g., ants, honey bees, vespid wasps, termites) provides a striking example of convergent phenotypic innovation (7). Eusocial insect societies are founded upon a novel caste polyphenism, in which reproductive queen and non-reproductive worker female castes develop from the same genome, depending mainly on socially-regulated nutritional inputs (8, 9). Within the worker caste, further specialization occurs as individuals age and progress through a series of tasks, including nursing and foraging (7). Polyphenic traits are often thought to evolve from pre-existing developmental plasticity (10), and leading hypotheses for the evolution of caste-based division of labor in social insects also stress the use and modification of highly conserved developmental and physiological mechanisms (11–15). The ovarian groundplan hypothesis (11, 16), subsequently elaborated as the reproductive groundplan hypothesis (13, 17), states that caste-based division of labor in eusocial lineages is derived from ancient reproductive physiological machinery, such that the reproductive and non-reproductive phases of the ancestral solitary life cycle have been decoupled to produce separate reproductive and non-reproductive castes. Similarly, the genetic toolkit hypothesis proposes that the convergent evolution of novel social behavior is driven by changing the regulation of a core "toolkit" of genes underlying highly conserved physiological processes such as metabolism (14, 15).

Studies focused on candidate genes underlying the genetic basis of caste-based division of labor within individual eusocial species have often found support for the importance of highly conserved genes and pathways associated with reproduction and metabolism. For example,

worker division of labor in honey bees is regulated by interactions between juvenile hormone, vitellogenin, and insulin/TOR signaling pathways (13, 17, 18), and similar pathways also play key roles in regulating division of labor between queen and worker castes in both ants and honey bees (19–22). While comparative genomic and transcriptomic studies have often similarly emphasized common general functions such as metabolism, such studies have thus far only identified very small sets of specific genes associated with the convergent evolution of caste, worker behavior, or eusociality in independent lineages (15, 23–27). Alternatively, many transcriptomic studies have argued for the importance of taxonomically-restricted genes for the evolution of caste-based division of labor (28–34). However, it is unclear if the lack of common specific genes is due to biological differences between the species or methodological details because studies in each species were not designed, conducted, or analyzed in parallel.

Previous studies have focused mainly on identifying whether there is significant overlap of genes or gene pathways associated with caste-based division of labor between independent lineages (24, 26, 27), but there has been little effort to quantify the relative importance of shared versus unshared genes to the convergent evolution of caste-based division of labor. Most of these studies have either focused on brain or whole body samples (15, 21, 23, 24, 34–37), although expression bias between queens and workers has been shown to vary strongly, depending on developmental stage and tissue type (29, 38–40). Surprisingly, no comparative study has investigated gene expression in queen and worker abdomens, the location of reproductive organs, where transcriptomic signatures of differences in reproductive physiology are the strongest (29). As such, while empirical studies in honey bees support the role of a reproductive groundplan in

regulating worker division of labor (13, 17), empirical studies have not yet addressed the role of a reproductive groundplan in the convergent evolution of reproductive caste.

Here, we present to date the most comprehensive developmental transcriptomic dataset investigating gene expression associated with reproductive caste and worker age-based division of labor in the pharaoh ant (*Monomorium pharaonis*) and the honey bee (*Apis mellifera*), species which represent two independent origins and elaborations of eusociality (41). We performed all sampling, sequencing, and analysis for the two species in parallel to maximize compatibility between the data sets. We leverage this extensive dataset to quantify in an unbiased manner the relative contribution of differential expression of shared versus distinct genes at each life stage and tissue to the convergent evolution of caste-based 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 we sequenced 177 RNA-sequencing libraries, across 28 distinct sample types for each species (Table S1).

Differential expression between queens and workers

To identify genes associated with caste development and adult caste dimorphism, we performed differential expression analysis between queens and workers at each developmental stage and adult tissue, separately for each species. The number of differentially expressed genes (DEGs) between queens and workers increased throughout development, peaking in the adult abdomen

(Fig. 1*A*). In all tissues and stages, the majority of caste-associated DEGs in one species were not differentially expressed in the other species (Fig. 1*A*). Similarly, the magnitude of gene-wise caste bias (as measured by $\log_2$ fold-change between queen and worker samples) was weakly 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 (Fig. S1; $r_{head} = 0.089$; $r_{thorax} = 0.161$; $r_{abdomen} = 0.275$; N = 7640 1:1 orthologs; N = 7460 1:1 orthologs; $P < 0.001$ in all cases). The top enriched Gene Ontology (GO) terms for caste-associated DEGs were dominated by metabolism, signaling and developmental processes (Table S3 and 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 each adult tissue, separately for each species. In general, there were very few behavioral DEGs in common in the two species (Fig. 1*B*). Gene-wise $\log_2$ fold-change between nurses and foragers was significantly but weakly correlated across ant and honey bee orthologs (Fig. S2; $r_{head} = 0.070$, $P_{head} < 0.001$; $r_{thorax} = 0.031$, $P_{thorax} = 0.008$; $r_{abdomen} = 0.051$, $P_{abdomen} < 0.001$; N = 7460 1:1 orthologs). The top enriched GO terms for behavioral DEGs were dominated by metabolism and developmental processes (Table S6 and S7).

Shared 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, approximately one third of abdominal caste-associated DEGs were common to both species (Fig. 1*A*; 1545 shared DEGs, comprising 35% [1545/4395] of ant DEGs, and 29% [1545/5352] of honey bee DEGs). Most shared abdominal differential expression was the result of shared queen-bias: 56% (858/1545 genes) of shared abdominal caste-associated DEGs 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). Shared abdominal caste-associated DEGs were more likely to be identified as evolutionarily ancient in comparison to non-biased genes (Fig. 1*C*; Fisher Test; $F = 3.41$, $P < 0.001$). Furthermore, abdominal DEGs with shared queen bias were more likely to be identified as ancient than DEGs with shared worker bias (Fig. 1*C*; Fisher Test; $F = 2.51$, $P < 0.001$). In general, genes varied in expression bias across evolutionary age categories, though the direction of the effect was not consistent across all tissues and stages (Fig. S3).

We next tried to put the seemingly large proportion of shared abdominal caste-associated DEGs (35% for ants and 29% for honey bees) into context, through comparison with the proportion of genes that were differentially expressed across larval development in both species, given that the molecular mechanisms of development are thought to be highly conserved (42). We identified 6089 and 6225 developmental DEGs in ants and honey bees, respectively, including 2544 shared DEGs, representing 42% (2544/6089) and 41% (2544/6255) of the total developmental DEGs in each species (Fig. S4).

To identify which of the thousands of abdominal DEGs found in each species are particularly important for queen abdominal expression (and presumably function), we performed gene co-expression analysis, separately for each species. We identified a module of genes specifically associated with queen abdominal expression in each species (N = 1006 genes in module for ants, N = 1174 genes for honey bees). We identified hub genes in each module, genes which are centrally connected in networks and strongly associated queen abdominal expression (43). Many hub genes were clearly associated with reproduction and maternal effects (Table S8 and S9), including genes with known roles in caste determination such as vitellogenin (*Vg receptor* was identified in each species) (22) and *vasa* (44), while others are important maternal proteins such as *Smaug* (45) and *ovo* (46). 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 were more centrally-located within modules than genes found in only one species-specific module (Fig. 2 *C* and *D*).

Caste bias is in part derived from ancestral sex bias

Given that our co-expression analysis indicated that many of the most important queen-upregulated 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 $\log_2$ fold change between queen and worker abdomens and gene-wise $\log_2$ fold-change between queen and male abdomens in both honey bees and pharaoh ants (Fig. 3 *A* and *B*). Additionally, sex bias itself was correlated between species (Fig. 3*C*). The correlation of

caste bias 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 (Fig. S5).

Given the link between shared caste bias and sex bias within ants and honey bees, we hypothesized that these shared caste-biased genes were derived from conserved pathways that also underlie sexual dimorphism in distant relatives. To test this hypothesis, we estimated the whole-body sex bias of orthologs in the fruit fly *Drosophila melanogaster* using available data (42). Shared queen-biased abdominal DEGs tended to be upregulated in females in *D. melanogaster* (Fig. 3*D*; one-sided Binomial Test for likelihood of queen conserved having $\log_2$ fold-change > 0; $P < 0.001$; N = 566 shared queen DEGs), while shared worker-biased abdominal DEGs tended to be upregulated in males ($P < 0.001$; N = 160 shared worker DEGs).

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, differential expression based on either reproductive caste or worker division of labor was largely not shared between species (Fig. 1 *A* and *B*). Furthermore, genes were often differentially expressed across many stages and tissues and sometimes in opposite directions (Fig. S6; e.g., upregulated in queen heads but downregulated in queen abdomens). To quantify the degree to which genes exhibited biased expression according to reproductive caste across all developmental stages and tissues, we calculated gene-wise “overall caste bias” in each species, where overall caste bias is the Euclidean distance of $\log_2$ fold-change across all queen/worker comparisons separately for each

species (see (47)). Similarly, we defined "overall behavior bias" as the Euclidean distance of $\log_2$ fold-change across all nurse/forager comparisons, separately for each species.

Across 1:1 orthologs, overall caste bias measured in ants was correlated to overall caste bias measured in honey bees, and overall behavior bias was similarly correlated between species (Fig. S7). Within species, overall caste and behavior bias were also correlated to each other (Fig. S8). 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 (Fig. 4 *A* and *B*) and behavior bias (Fig. S9 *A* and *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 (i.e. peripheral network elements) in co-expression networks constructed across all samples tended to exhibit more caste and behavior bias in comparison to highly connected genes (Fig. 4 *C* and *D*; Fig. S9 *C* and *D*). 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 (Fig. S10), where tissue specificity was calculated using available data (31). Finally, genes that were rapidly evolving (i.e. with high values of dN/dS) tended to exhibit higher levels of caste and behavior bias (Fig. 4 *E* and *F*; Fig. S9 *E* and *F*). Importantly, while expression is correlated to overall caste and behavior bias, our results are not driven by the effect of expression levels according to partial correlation analysis (Table S11).

Discussion

The reproductive groundplan hypothesis proposes that the convergent evolution of caste-based division of labor is derived from reproductive-associated pathways found in the long-extinct solitary ancestors of social insects (11, 13, 16, 17). In strong support of this hypothesis, we identify a large set (~1500) of genes with shared caste-biased abdominal expression in pharaoh ants and honey bees (Fig. 1*A*), including many genes with known roles in reproduction such as the *vitellogenin* receptor (22) and *ovo* (46). This set of shared caste-biased genes seems to be derived from ancient plastically-expressed genes underlying sexual dimorphism, as genes upregulated in queen abdomens of both ants and honey bees tend to also be female-biased in the distant relative *Drosophila melanogaster* (Fig. 3*D*). Previous studies had failed to find large sets of genes repeatedly used for eusocial evolution (23–27), but no previous comparative study investigated caste-biased expression in the abdomen. The large overlap for abdominal caste-associated genes is notable because honey bees and ants last shared a common ancestor approximately 160 million years ago (48), and this overlap is nearly as much as we see for genes that were differentially expressed across developmental stages (Fig. S4). Shared developmental

molecular mechanisms are presumably due simply to shared ancestry and the deep conservation of developmental mechanisms (42, 49). The similar level of overlap for caste-associated genes points to the large-scale recruitment of pre-existing developmental and physiological machinery underlying the ancestral reproductive groundplan during the independent evolution of caste-based division of labor in ant and honey bee lineages.

While reproductive caste in complex eusocial societies such as ants and honey bees is typically fixed in adulthood, the tasks performed by workers (specifically, nursing versus foraging) change over the course of the worker's adult lifetime (18, 34). This plastic 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 (18, 20). However, we identified few genes that were commonly differentially expressed between nurses and foragers in honey bees and pharaoh ants (Fig. 1*B*), and 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 enriched Gene Ontology categories associated with development as well as metabolism in each species (Tables S6 and 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 (14, 24, 27).

Conserved factors or pathways clearly play important roles in aspects of caste development and function as well as the transition from nursing to foraging, but 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 (24, 30, 34–36). This lineage-specific

architecture is comprised of large groups of both orthologous genes with different expression patterns and taxonomically-restricted genes (Fig. 1 *A* and *B*). In contrast the low amount of context-specific overlap in differential expression, the overall degree of caste-associated plastic expression across stages and tissues (overall caste bias) was correlated between species (Fig. S7 *A* and *B*), and expression plasticity was correlated between behavior and caste (Fig. S8). Genes with high overall caste bias tended to be loosely connected (Fig. 4 *C* and *D*, i.e. downstream in regulatory networks) and displayed tissue-specific expression profiles (Fig. S10). This indicates that changes in their expression are unlikely to strongly influence the expression of many other genes. Furthermore, we find that genes with high overall caste bias are weakly constrained in terms of sequence evolution, as they tend to evolve rapidly and be evolutionarily young (Fig. 4 *A, B, E, F*), though note that evolutionary age and rate cannot be reliably disentangled (50). Evolutionary developmental biology emphasizes the action of a certain type of gene for the evolution of morphological novelty: highly connected core elements of that are highly pleiotropic and slowly evolving (49, 51). However, the emerging picture is that the evolution of many traits involve a majority of genes on the opposite end of the spectrum: genes with highly variable expression across contexts that are peripheral elements of regulatory networks and exhibit high levels of divergence in expression and sequence between species (31, 34, 52).

Conclusions

Our study shows that the recruitment of a large core of conserved reproductive-associated genes, which make up a reproductive groundplan, is fundamental to the convergent evolution of caste-based division of labor in ants and honey bees. However, our study also reveals that the

bulk of the full genetic architecture underlying the expression of social insect caste-based division of labor varies between lineages. This is reflected by the general biology of social insects, in that independently evolved societies share reproductive division of labor, the main defining feature of eusociality, but also display a wide diversity of lineage-specific adaptations (7). 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 is a myriad of taxonomically-restricted genes as well as conserved genes with lineage-specific expression patterns (6, 31, 34, 53). This is consistent with models for the evolution of hierarchical developmental gene regulatory networks, whereby a relatively small number of highly conserved genes act upstream to initiate gene cascades (e.g., to set up body-patterning), while batteries of downstream genes are evolutionarily labile and largely responsible for lineage-specific features (54). Recent studies have made progress elucidating the function of several core genes and pathways for caste (19–21, 55). Large-scale transcriptomic studies such as ours serve a complimentary, indispensable role of identifying the full suite of genes underlying caste-based division of labor in multiple independent lineages.

Methods

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 (head, thorax, abdomen) in both species (Table S1).

Differential Expression Analysis

To identify caste-associated differentially-expressed genes (DEGs), we performed differential expression analysis between queens and workers at each developmental stage and tissue, separately for each species. We first removed lowly-expressed genes with genes with counts per million (CPM) less than one in all samples. We constructed GLM-like models including replicate and caste and identified genes associated with caste (FDR < 0.1) at each stage or tissue using EdgeR (56). Similarly, to identify behavioral DEGs we performed differential expression analysis between nurses and foragers for each tissue. To identify developmental DEGs in each species, we constructed models with all larval and egg samples and identified genes differentially expressed between any developmental stage (FDR < 0.1), controlling for overall caste differences. To estimate gene-wise sex-bias of *D. melanogaster* orthologs, we downloaded available whole body RNA-seq data (42), consisting of one 5-day and one 30-day fly of each sex, and performed differential expression analysis as above.

Coexpression Network Analysis

We performed plaid clustering, a non-deterministic biclustering algorithm (57). Biclustering seeks to identify groups of genes that are co-expressed across a specific subset of samples (58). We identified genes that were consistently associated with a queen-abdomen specific bicluster across 1000 iterations, which we term queen abdominal modules (see Supplemental Methods). We conservatively identified module hub genes as genes with intra-module connectivity in at least the 90th percentile and abdominal $\log_2$ fold-change values greater than 2 (representing a 4-fold increase in expression in queen relative to worker abdomens). We calculated connectivity of

each gene as the weighted sum of the Pearson correlation of expression between the given gene and all other genes, where we raised each correlation to the 6th power, the default value for weighted gene co-expression analysis (59). Intra-module connectivity (used in Fig. 2) represents the connectivity of genes within the queen abdominal module to other genes within the module, while total network connectivity (used in Fig. 4) represents the connectivity of genes across the entire transcriptome, after filtering out lowly-expressed genes.

Estimation of tissue specificity

We downloaded available RNA-sequencing on twelve tissues in *A. mellifera* in worker nurses and foragers (31). To classify genes by their tissue specificity, we calculated τ, a commonly used metric of expression specificity (60). τ ranges from 0 to 1, where 0 indicates genes are ubiquitously expressed and 1 indicates genes are exclusively expressed in one tissue.

Estimation of evolutionary rate

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 with an available genome (*A. mellifera*: *A. cerana*; *M. pharaonis*: *S. invicta*). 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 (61), derived nucleotide alignments from protein alignments using pal2nal (62), and estimated pairwise dN/dS of nucleotide alignments using PAML, package codeml (63).

Partial correlation analysis

We performed partial Spearman correlations between overall bias and evolutionary/network characteristics, controlling for the effect of expression.

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 would like to thank the following: Luigi Pontieri for photographs of pharaoh ants, Alex Wild for photographs of honey bees, Chao Tong for compiling hymenopteran genomes for use in phylostratigraphy and for comments on the manuscript, Rohini Singh for comments on the manuscript, and Junhyong Kim, Mia Levine, and Justin Walsh for helpful comments and discussion. This work was funded by the National Science Foundation (grant number IOS-1452520 to TAL) and 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. The National Science Foundation also funded MRW (DGE-1321851).

Works Cited

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

2. Conte GL, Arnegard ME, Peichel CL, Schluter D (2012) The probability of genetic parallelism and convergence in natural populations. *Proc Biol Sci* 279(1749):5039–5047.

3. Zhao L, Begun DJ (2017) Genomics of parallel adaptation at two timescales in *Drosophila*. *PLoS Genet* 13(10):e1007016.

4. Shubin N, Tabin C, Carroll S (2009) Deep homology and the origins of evolutionary novelty. *Nature* 457(7231):818–823.

5. Wilson GA, et al. (2005) Orphans as taxonomically restricted and ecologically important genes. *Microbiology* 151(Pt 8):2499–2501.

6. Johnson BR (2018) Taxonomically restricted genes are fundamental to biology and evolution. *Front Genet* 9:407.

7. Wilson EO (1971) *The insect societies*. (Harvard University Press, Boston).

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

9. Linksvayer TA, et al. (2011) Larval and nurse worker control of developmental plasticity and the evolution of honey bee queen–worker dimorphism. *J Evol Biol* 24(9):1939–1948.

10. Nijhout HF (2003) Development and evolution of adaptive polyphenisms. *Evol Dev* 5(1):9–18.

11. West-Eberhard, MJ. 1987. Flexible strategy and social evolution. *Animal societies Theories and facts*, eds Ito Y, Brown JL, Kikkawa, J (Scientific Societies Press, Tokyo), pp 35-51.

12. Linksvayer TA, Wade MJ (2005) The evolutionary origin and elaboration of sociality in the aculeate Hymenoptera: maternal effects, sib-social effects, and heterochrony. *Q Rev Biol* 80(3):317–336.

13. Amdam GV, Csondes A, Fondrk MK, Page RE Jr (2006) Complex social behaviour derived from maternal reproductive traits. *Nature* 439(7072):76–78.

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

15. Toth AL, et al. (2010) Brain transcriptomic analysis in paper wasps identifies genes associated with behaviour across social insect lineages. *Proc Biol Sci* 277(1691):2139–2148.

16. West-Eberhard MJ (1996) *Natural history and evolution of paper-wasps* (Oxford University Press, Oxford).

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

18. Robinson GE (1987) Regulation of honey bee age polyethism by juvenile hormone. *Behav Ecol Sociobiol* 20(5):329–338.

19. Mutti NS, et al. (2011) IRS and TOR nutrient-signaling pathways act via juvenile hormone to influence honey bee caste fate. *J Exp Biol* 214(Pt 23):3977–3984.

20. Gospocic J, et al. (2017) The Neuropeptide corazonin controls social behavior and caste identity in ants. *Cell* 170(4):748–759.e12.

21. Chandra V, et al. (2018) Social regulation of insulin signaling and the evolution of eusociality in ants. *Science* 361(6400):398–402.

22. Libbrecht R, et al. (2013) Interplay between insulin signaling, juvenile hormone, and vitellogenin regulates maternal effects on polyphenism in ants. *Proc Natl Acad Sci U S A* 110(27):11050–11055.

23. Toth AL, et al. (2014) Shared genes related to aggression, rather than chemical communication, are associated with reproductive dominance in paper wasps (*Polistes metricus*). *BMC Genomics* 15:75.

24. Berens AJ, Hunt JH, Toth AL (2015) Comparative transcriptomics of convergent evolution: different genes but conserved pathways underlie caste phenotypes across lineages of eusocial insects. *Mol Biol Evol* 32(3):690–703.

25. Woodard SH, Bloch GM, Band MR, Robinson GE (2014) Molecular heterochrony and the evolution of sociality in bumblebees (*Bombus terrestris*). *Proc Biol Sci* 281(1780):20132419.

26. Woodard SH, et al. (2011) Genes involved in convergent evolution of eusociality in bees. *Proc Natl Acad Sci U S A* 108(18):7472–7477.

27. Kapheim KM, et al. (2015) Social evolution. Genomic signatures of evolutionary transitions from solitary to group living. *Science* 348(6239):1139–1143.

28. Feldmeyer B, Elsner D, Foitzik S (2014) Gene expression patterns associated with caste and reproductive status in ants: worker-specific genes are more derived than queen-specific ones. *Mol Ecol* 23(1):151–161.

29. Warner MR, Mikheyev AS, Linksvayer TA (2017) Genomic signature of kin selection in an

ant with obligately sterile workers. *Mol Biol Evol* 34(7):1780–1787.

30. Ferreira PG, et al. (2013) Transcriptome analyses of primitively eusocial wasps reveal novel insights into the evolution of sociality and the origin of alternative phenotypes. *Genome Biol* 14(2):R20.

31. Jasper WC, et al. (2015) Large-scale coding sequence change underlies the evolution of postdevelopmental novelty in honey bees. *Mol Biol Evol* 32(2):334–346.

32. Johnson BR, Tsutsui ND (2011) Taxonomically restricted genes are associated with the evolution of sociality in the honey bee. *BMC Genomics* 12:164.

33. Harpur BA, et al. (2014) Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. *Proc Natl Acad Sci U S A* 111(7):2614–2619.

34. Mikheyev AS, Linksvayer TA (2015) Genes associated with ant social behavior show distinct transcriptional and evolutionary patterns. *Elife* 4:e04775.

35. Morandin C, et al. (2016) Comparative transcriptomics reveals the conserved building blocks involved in parallel evolution of diverse phenotypic traits in ants. *Genome Biol* 17:43.

36. Qiu B, et al. (2018) Towards reconstructing the ancestral brain gene-network regulating caste differentiation in ants. *Nat Ecol Evol* 2(11):1782–1791.

37. Patalano S, et al. (2015) Molecular signatures of plastic phenotypes in two eusocial insect species with simple societies. *Proc Natl Acad Sci U S A* 112(45):13970–13975.

38. Morandin C, et al. (2015) Caste-biases in gene expression are specific to developmental stage in the ant *Formica exsecta*. *J Evol Biol* 28(9):1705–1718.

39. Ometto L, Shoemaker D, Ross KG, Keller L (2011) Evolution of gene expression in fire ants: the effects of developmental stage, caste, and species. *Mol Biol Evol* 28(4):1381–1392.

40. Lucas ER, Romiguier J, Keller L (2017) Gene expression is more strongly influenced by age than caste in the ant *Lasius niger*. *Mol Ecol* 26(19):5058–5073.

41. Branstetter MG, et al. (2017) Phylogenomic insights into the evolution of stinging wasps and the origins of ants and bees. *Curr Biol* 27(7):1019–1025.

42. Gerstein MB, et al. (2014) Comparative analysis of the transcriptome across distant species. *Nature* 512(7515):445–448.

43. Zhang B, Horvath S (2005) A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* 4:Article17.

44. Khila A, Abouheif E (2010) Evaluating the role of reproductive constraints in ant social

evolution. *Philos Trans R Soc Lond B Biol Sci* 365(1540):617–630.

45. Benoit B, et al. (2009) An essential role for the RNA-binding protein *Smaug* during the *Drosophila* maternal-to-zygotic transition. *Development* 136(6):923–932.

46. Mével-Ninio M, Terracol R, Salles C, Vincent A, Payre F (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. *Mech Dev* 49(1):83–95.

47. Schrader L, Helanterä H, Oettler J (2017) Accelerated evolution of developmentally biased genes in the tetraphenic ant *Cardiocondyla obscurior*. *Mol Biol Evol* 34(3):535–544.

48. Peters RS, et al. (2017) Evolutionary history of the Hymenoptera. *Curr Biol* 27(7):1013–1018.

49. Carroll SB (2008) Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134(1):25–36.

50. Moyers BA, Zhang J (2015) Phylostratigraphic bias creates spurious patterns of genome evolution. *Mol Biol Evol* 32(1):258–267.

51. Hoekstra HE, Coyne JA (2007) The locus of evolution: evo devo and the genetics of adaptation. *Evolution* 61(5):995–1016.

52. Boyle EA, Li YI, Pritchard JK (2017) An expanded view of complex traits: from polygenic to omnigenic. *Cell* 169(7):1177–1186.

53. Trible W, Kronauer DJC (2017) Caste development and evolution in ants: it's all about size. *J Exp Biol* 220(Pt 1):53–62.

54. Erwin DH, Davidson EH (2009) The evolution of hierarchical gene regulatory networks. *Nat Rev Genet* 10(2):141–148.

55. Simola DF, et al. (2016) Epigenetic (re)programming of caste-specific behavior in the ant *Camponotus floridanus*. *Science* 351(6268):aac6633.

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

57. Lazzeroni L, Owen A (2002) Plaid models for gene expression data. *Stat Sin* 12(1):61–86.

58. Tanay A, Sharan R, Shamir R (2002) Discovering statistically significant biclusters in gene expression data. *Bioinformatics* 18 Suppl 1:S136–44.

59. Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network

analysis. *BMC Bioinformatics* 9:559.

60. Yanai I, et al. (2005) Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* 21(5):650–659.

61. Larkin MA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23(21):2947–2948.

62. Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* 34(Web Server issue):W609–12.

63. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24(8):1586–1591.

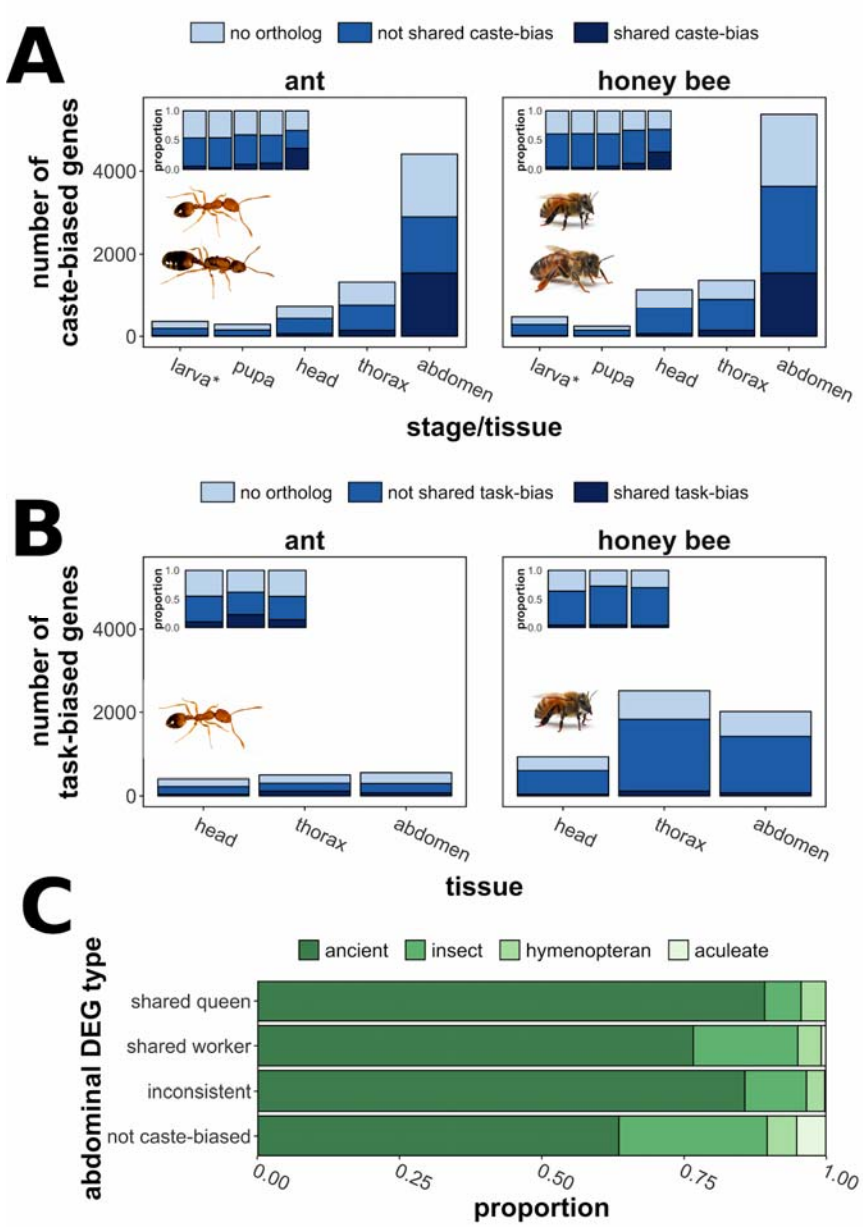

Fig. 1. Number of differentially expressed genes (FDR < 0.1) between A) queens and workers and B) nurses and foragers at each developmental stage or tissue in ants (left) and honey bees (right). "Head", "thorax", and "abdomen" refer to body segments of adults, while "pupa" and "larva" refer to whole bodies. "No ortholog" refers to genes for which no 1:1 ortholog exists (either due to apparent duplication or complete lack or orthology), "not shared caste/task bias" refers to genes for which 1:1 orthologs can be identified but are only differentially expressed in one species, and "shared caste/task" bias refers to genes for which 1:1 orthologs are differentially expressed in both species. Insets show the proportion of each category of gene out of all differentially expressed genes at that stage or tissue. C) Proportion of abdominal DEGs by estimated evolutionary age (shading). "Shared queen/worker" indicates genes upregulated in queen or workers of both species.

*****: the category "larva" represents differential expression across larvae of all stages for which caste can be identified (second to fifth larval stage).

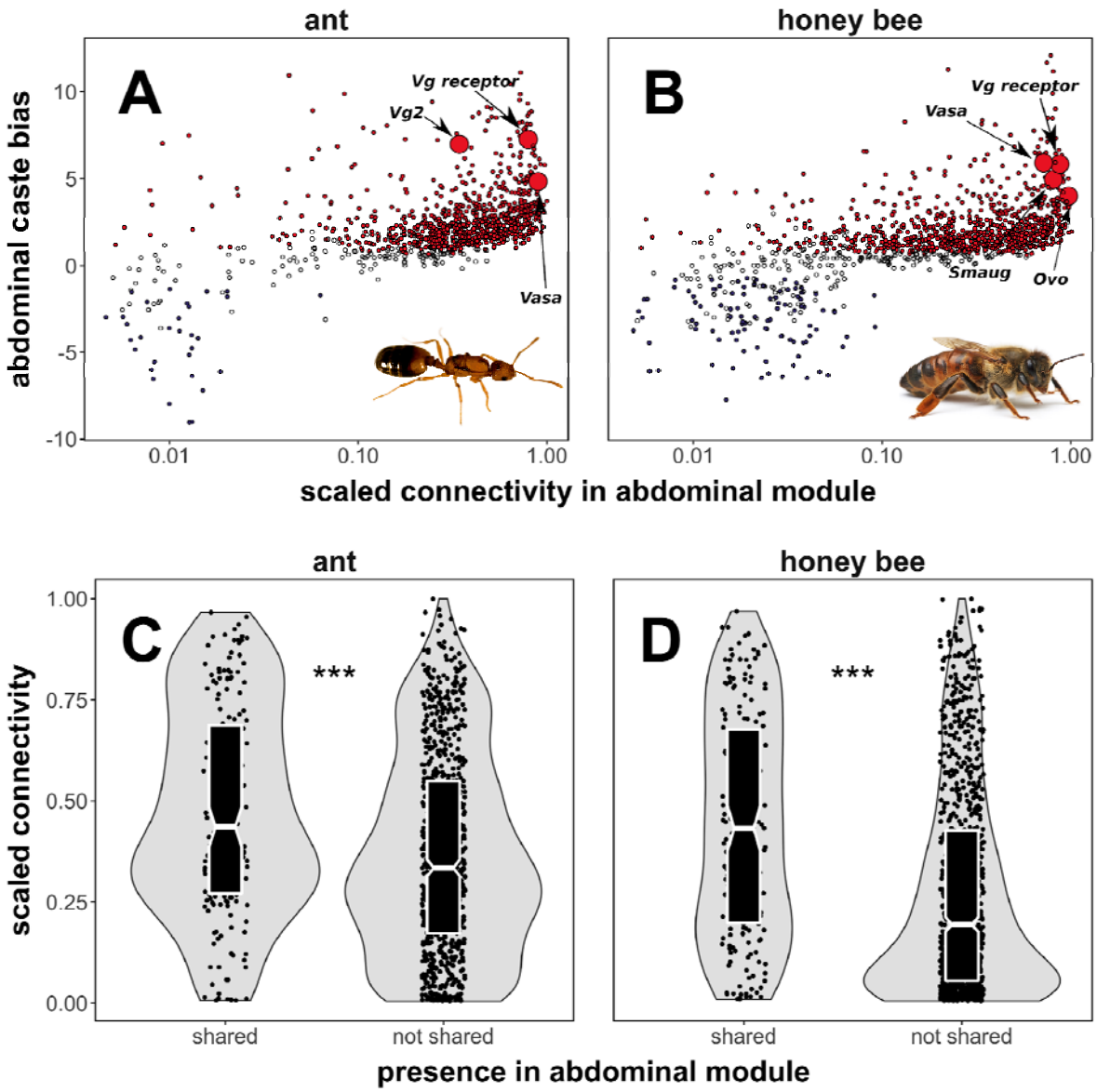

Fig. 2. Abdominal caste bias ($\log_2$ fold change queen versus worker) is correlated with connectivity within the queen-abdomen module in A) ants (rho = 0.536, P < 0.001) and B) honey bees (rho = 0.617, P < 0.001). Genes upregulated in queens are in red, while genes upregulated in workers are in blue. Connectivity is proportional to the most highly connected gene in the module. Connectivity within the queen abdominal module is higher for 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)

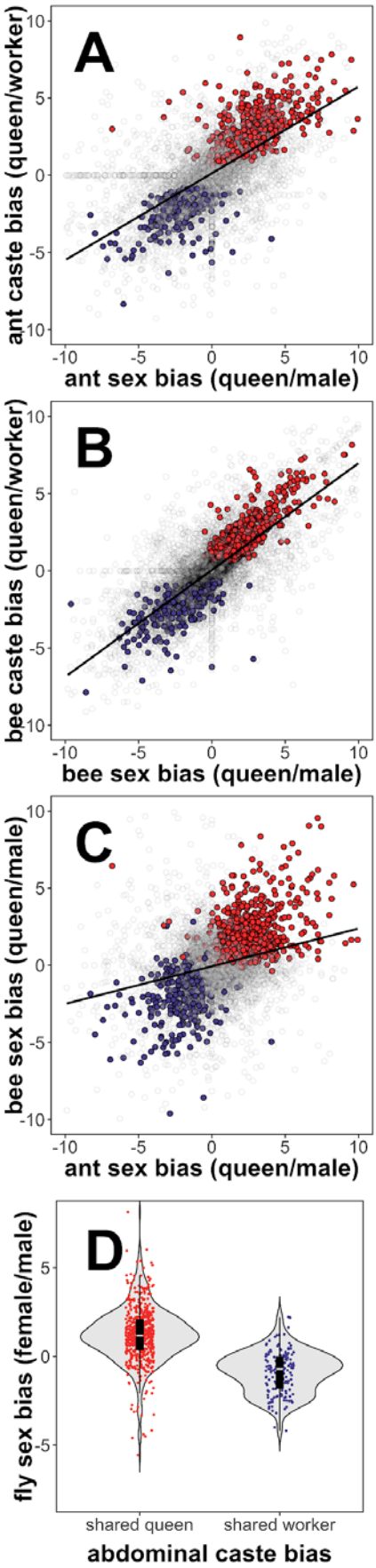

Fig. 3. Caste bias is derived from sex bias. Abdominal caste bias (queen vs worker $\log_2$ fold-change) is correlated to abdominal sex bias (queen vs male $\log_2$ fold-change) in A) *M. pharaonis* (rho = 0.715, P < 0.001) and B) *A. mellifera* (rho = 0.774, P < 0.001) and abdominal sex bias is correlated between the two species (rho = 0.280, P < 0.001) (C). Red indicates shared queen-biased abdominal DEGs, while blue indicates shared worker-biased abdominal DEGs. Grey indicates genes that did not exhibit shared expression patterns or were not differentially expressed. D) Shared queen-biased abdominal DEGs tend to be female-biased in *D. melanogaster* while shared worker-biased abdominal DEGs tend to be male-biased in *D. melanogaster* (likely reflecting down-regulation in females).

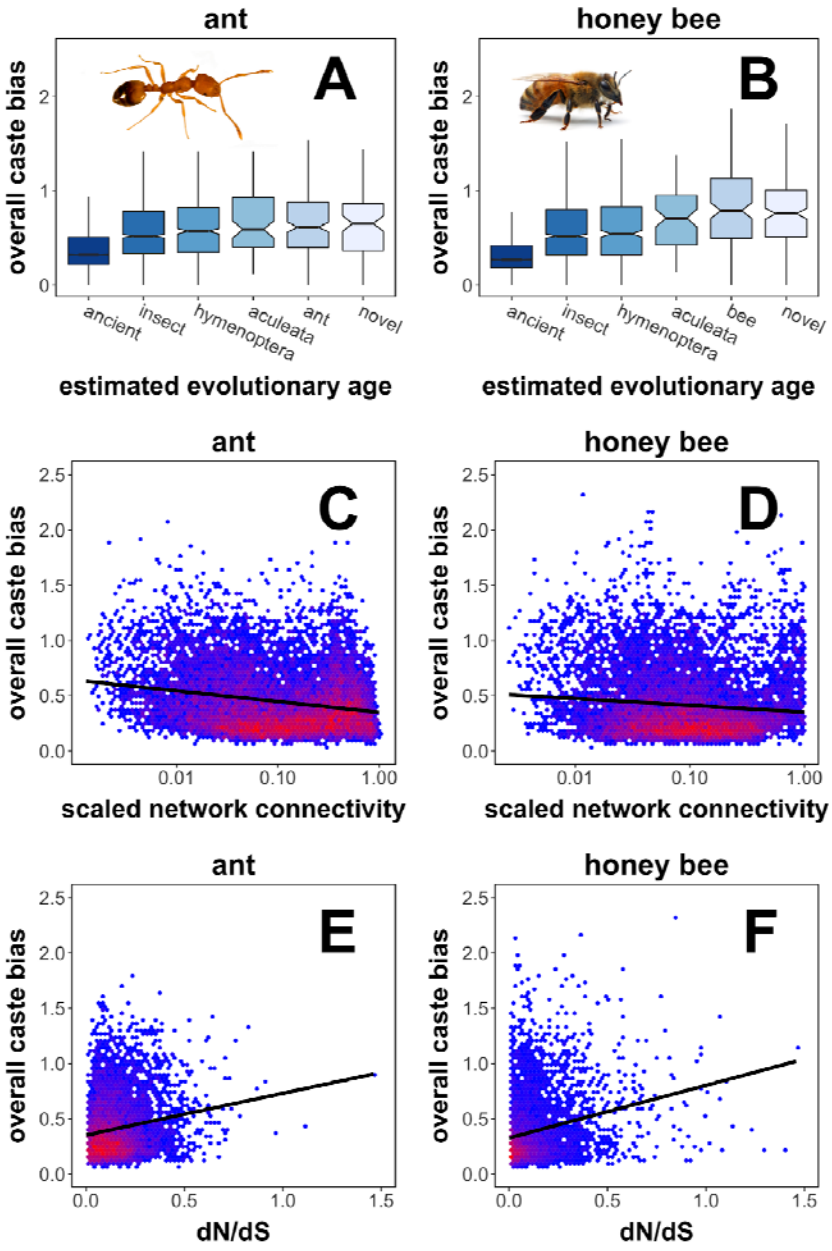

Fig. 4. Genes that exhibit more caste bias across tissues and developmental stages have younger estimated evolutionary ages (A,B) and tend to be loosely connected (C,D; ant: rho = -0.159, P < 0.001; honey bee: rho = -0.090, P < 0.001) and rapidly evolving (E,F; ant: rho = 0.157, P < 0.001; honey bee: rho = 0.240, P < 0.001). "Overall caste bias" combines queen/worker $\log_2$ 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.