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 is also 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.

Introduction

The degree to which convergent phenotypic evolution involves the same sets of genes or pathways is a major unanswered question (1, 2). Comparative genomic studies indicate that parallel adaptive genetic 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 (3, 4), but convergent adaptive changes have been implicated only in small subsets of genes across distantly related species (5, 6, see 7, 8). 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 (9, 10), as exemplified by the important role the transcription factor *Pax6* plays in eye development in molluscs, vertebrates, and insects, which each evolved eyes independently (11, 12). 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 (13) have been shown to be important for lineage-specific evolutionary novelties such as cnidarian nematocysts (14, 15), 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 (16). These eusocial 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 (17, 18). Within the worker caste, further specialization occurs as individuals age and progress through a series of tasks, including nursing and foraging

51 (19, 20). Polyphenic traits are often thought to evolve from pre-existing developmental plasticity
52 (21), and leading hypotheses for the evolution of caste-based division of labor in social insects
53 also stress the use and modification of highly conserved developmental and physiological
54 mechanisms (22–26). The ovarian groundplan hypothesis (22, 27), subsequently elaborated as
55 the reproductive groundplan hypothesis (24, 28) states that the derived caste-based division of
56 labor in eusocial lineages is constructed from ancient reproductive physiological machinery, such
57 that the reproductive and non-reproductive phases of the ancestral life cycle have been decoupled
58 to produce separate reproductive and non-reproductive castes. Similarly, the genetic toolkit
59 hypothesis proposes that the convergent evolution of novel social behavior is driven by changing
60 the regulation of a core "toolkit" of genes underlying highly conserved physiological processes
61 such as metabolism (25, 26).

62 Studies focused on candidate genes underlying the genetic basis of caste-based division
63 of labor within individual eusocial species have often found support for the importance of highly
64 conserved genes and pathways associated with reproduction and metabolism. For example,
65 worker division of labor in honey bees is regulated by interactions between juvenile hormone,
66 vitellogenin, and insulin/TOR signaling pathways (24, 28–30), and similar pathways also play
67 key roles in regulating division of labor between queen and worker castes in both honey bees and
68 ants (31–36). While comparative genomic and transcriptomic studies have often similarly
69 emphasized common general functions such as metabolism, such studies have thus far only
70 identified very small sets of specific genes associated with the convergent evolution of caste,
71 worker behavior, or eusociality in independent lineages (26, 37–42). Alternatively, many
72 transcriptomic studies have argued for the importance of taxonomically-restricted genes for the

evolution of caste-based division of labor, for example genes lacking orthologs in separate eusocial lineages (43–49). 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 (38, 41, 42), 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 (26, 34, 37–39, 49–51), although expression bias between queens and workers has been shown to vary strongly, depending on developmental stage and tissue type (44, 52–54). Surprisingly, only two studies (44, 55), each in a single species, compared gene expression in queen and worker abdomens, the location of reproductive organs, where transcriptomic signatures of differences in reproductive physiology are the strongest (44). As such, while empirical studies in honey bees support the role of a reproductive groundplan in regulating worker division of labor (24, 28), 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 honey bee (*Apis mellifera*) and the pharaoh ant (*Monomorium pharaonis*), species which represent two independent origins and elaborations of eusociality (56). 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 quantify in an unbiased manner the relative contribution of differential expression at shared genes versus distinct genes to the convergent evolution of reproductive caste and worker division of labor, and we also assess the relative importance of highly conserved versus taxonomically-restricted genes at each life stage and tissue to the expression of these traits. In support of the reproductive groundplan hypothesis, we identify a large core set of reproduction-related genes upregulated in ant and honey bee queen abdomens, which is derived from a set of ancient sex-biased genes. However, this shared core comprises only one third of total abdominal differential expression, and outside the abdomen, distinct sets of genes 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 we sequenced 177 RNA-sequencing libraries across 28 distinct sample types (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 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, 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. The number of behavioral (i.e. between nurses and foragers) DEGs for a given tissue varied from 405 in ant heads to 2519 in honey bee thoraces (Table S5). There were more behavioral DEGs in honey bees than in ants for all three adult tissues. 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).

139

140 *Shared abdominal caste-bias in ancient genes*

141 For the most part our results indicate distinct genes are associated with caste and worker division
142 of labor in honey bees and ants. However, approximately one third of abdominal caste-associated
143 DEGs were common to both species (Fig. 1*A*; 1545 shared DEGs, comprising 35% [1545/4395]
144 of ant DEGs, and 29% [1545/5352] of honey bee DEGs). Most shared abdominal differential
145 expression was the result of shared queen-bias: 56% (858/1545 genes) of shared abdominal
146 caste-associated DEGs were upregulated in queen abdomens in both species, compared to 22%
147 (338/1545) that were worker-upregulated and 23% (349/1545) that reversed direction (i.e. were
148 queen-biased in one species and worker-biased in the other). Shared abdominal caste-associated
149 DEGs were more likely to be identified as evolutionarily ancient in comparison to non-biased
150 genes (Fig. 1*C*; Fisher Test; $F = 3.41$, $P < 0.001$). Furthermore, abdominal DEGs with shared
151 queen bias were more likely to be identified as ancient than DEGs with shared worker bias (Fig.
152 1*C*; Fisher Test; $F = 2.51$, $P < 0.001$), though the estimated evolutionary age of genes was not
153 consistently associated with queen or worker bias across all tissues and stages (Fig. S3).

154 We next tried to put the seemingly large amount of overlap between our species for
155 abdominal caste-associated DEGs (1545 shared DEGs, ~⅓ of total DEGs) into context, since it is
156 unclear how much overlap to expect based on the fact that pharaoh ants and honey bees share
157 developmental and physiological pathways due to common descent (the last common ancestor
158 existed ~160 million years ago) (57). We first calculated the genome-wide proportion of genes
159 which had 1:1 orthologs in each species and were expressed at functionally-relevant levels. This
160 serves as a null expectation for the proportion of genes that are common by descent and might be

161 used for convergent phenotypic evolution, if these genes were recruited at random from the
162 genome. Out of genes with counts per million (CPM) greater than one in at least one sample
163 type, 69% (7460/10804) of genes in ants and 63% (7460/11775) of genes in honey bees had a
164 corresponding 1:1 ortholog in the other species. These proportions correspond roughly with the
165 proportion of caste-associated DEGs that have 1:1 orthologs overall (Fig. 1 *A* and *B*), suggesting
166 that the transcriptomic architecture of caste has a similar proportion of shared genes (i.e.
167 common by descent) as the rest of the genome. Next, we identified genes that were differentially
168 expressed across larval development to compare the overall make-up of developmental DEGs to
169 caste-associated DEGs, given that the molecular mechanisms of development are thought to be
170 highly conserved (58). We identified 6089 developmental DEGs in ants and 6225 developmental
171 DEGs in honey bees (Fig. S4). 68% (4133/6089) and 70% (4327/6225) of these developmental
172 DEGs in ants and honey bees, respectively, had 1:1 orthologs, and 2544 developmental DEGs
173 were shared between species, representing 42% (2544/6089) and 41% (2544/6255) of the total
174 developmental DEGs in each species (Fig. S4). While a smaller proportion of caste-associated
175 DEGs (35% for ants and 29% for honey bees) were shared between species, it is notable that
176 these proportions are nearly as high as those we observe for development.
177
178 *Core queen-biased abdominal genes are associated with reproduction*
179 To identify core genes associated with queen abdominal expression, we performed gene co-
180 expression analysis, separately for each species. We focused on identifying a group of genes (i.e.
181 module) which exhibited correlated expression patterns specifically across queen abdominal
182 samples (see Methods). We identified a queen abdominal module of 1006 genes in ants and a

183 module of 1174 genes in honey bees. We calculated the connectivity of each gene within these
184 identified abdominal modules (i.e. intra-modular connectivity), where genes with higher
185 connectivity values are central to the module, i.e. they exhibit highly correlated expression
186 patterns with many other genes in the module. There was a positive correlation between intra-
187 modular connectivity and $\log_2$ fold-change between queen and worker abdomens (Fig. 2 *A* and
188 *B*; ants: rho = 0.536, P < 0.001, N = 1006; honey bees: rho = 0.617, P < 0.001, N = 1174), as
189 well a positive correlation between intra-modular connectivity and the significance ($-\log_{10}$ of P-
190 value) of queen/worker abdominal differential expression (Spearman correlation; ants: rho =
191 0.485, P < 0.001, N = 1006; honey bees: rho = 0.555, P < 0.001, N = 1174), indicating that the
192 most tightly connected genes within these modules show strong queen-biased expression. We
193 identified 183 genes which were present in the queen abdominal module of both species, and
194 these genes were overwhelmingly queen-biased (78.7% [144/183] upregulated in queens of both
195 species), and had higher intra-module connectivity than genes found in only one species-specific
196 module (Fig. 2 *C* and *D*).
197 A major strength of gene co-expression analysis is its power to identify hub genes for
198 traits, i.e. genes which are centrally connected and strongly associated with a trait (59). We
199 identified hub genes in both species (see Methods), many of which are clearly associated with
200 reproduction and maternal effects (Table S8 and S9; Fig. 2 *A* and *B*). Included in this list are
201 genes with previous known roles in caste and reproduction such as vitellogenin (*Vg receptor* was
202 identified in each species) (31, 60) and *vasa* (61), while others are important maternal proteins
203 such as *Smaug* (62) and *ovo* (63).
204

205 *Caste bias is in part derived from ancestral sex bias*

206 Given that our co-expression analysis indicated that many of the most important queen-
207 upregulated genes are clearly associated with female reproduction, we reasoned that caste-biased
208 expression may be derived from sex-biased expression. Indeed, there was a positive correlation
209 between gene-wise $\log_2$ fold change between queen and worker abdomens and gene-wise $\log_2$
210 fold-change between queen and male abdomens in both honey bees and pharaoh ants (Fig. 3 *A*
211 and *B*). Additionally, sex bias itself was correlated between species (Fig. 3*C*). The correlation of
212 caste bias and sex bias was not restricted to the abdomen, as there was similar highly significant
213 effect when comparing heads and thoraces, albeit with weaker effect sizes (Fig. S5).

214 Given the link between shared caste bias and sex bias within ants and honey bees, we
215 hypothesized that caste bias is derived from ancient pathways underlying sexual dimorphism. To
216 test this hypothesis, we estimated the whole-body sex-bias of orthologs in the fruit fly
217 *Drosophila melanogaster* using available data (58). Shared queen-biased abdominal DEGs
218 tended to be upregulated in females in *D. melanogaster* (Fig. 3*D*; one-sided Binomial Test for
219 likelihood of queen conserved having $\log_2$ fold-change > 0; $P < 0.001$; N = 566 shared queen
220 DEGs), while shared worker-biased abdominal DEGs tended to be upregulated in males ($P <$
221 0.001; N = 160 shared worker DEGs).

222

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

224 While we have emphasized the conservation of abdominal differential expression between
225 queens and workers in pharaoh ants and honey bees, roughly two-thirds of abdominal caste-
226 biased genes were not shared between species, and differential expression based on reproductive

227 caste or worker division of labor was largely not shared (Fig. 1 *A* and *B*). Furthermore, genes
228 were often differentially expressed across many stages and tissues and sometimes in opposite
229 directions (Fig. S6; e.g., upregulated in queen heads but downregulated in queen abdomens). To
230 quantify the degree to which genes exhibited biased expression according to reproductive caste
231 across all developmental stages and tissues we calculated the Euclidean distance of $\log_2$ fold-
232 change across all queen/worker comparisons separately for each species, and we labeled this
233 quantity "overall caste bias". This approach is analogous to previous work on expression
234 plasticity in ants across alternative morphs (64). Similarly, we defined "overall behavior bias" as
235 the Euclidean distance of $\log_2$ fold-change across all nurse/forager comparisons, separately for
236 each species.

237 Interestingly, both overall caste and overall behavior bias were correlated among
238 orthologs between species (Fig. S7; i.e. overall caste bias measured in ants was correlated with
239 overall caste bias measured in honey bees across 1:1 orthologs). Within species, overall caste and
240 behavior bias were also correlated to each other (Fig. S8). This indicates that plasticity in gene
241 expression is correlated across contexts (caste versus behavior) and species. GO terms associated
242 with high caste bias were largely linked to metabolism, while those associated with high
243 behavior bias were largely linked to developmental processes (Table S10).

244

245 *Characteristics of genes associated with caste and behavior*

246 We compared overall caste bias and overall behavior bias to gene age, evolutionary rate, network
247 connectivity, and tissue-specificity to understand the general features of genes commonly
248 associated with caste (queen versus worker) or behavior (nursing versus foraging). Genes with

249 younger estimated evolutionary ages tended to exhibit higher overall caste bias (Fig. 4 *A* and *B*)
250 and behavior bias (Fig. S9 *A* and *B*) compared in particular to ancient genes (Gamma GLM; ant
251 caste bias: $\chi^2 = 900.19$, honey bee caste bias: $\chi^2 = 1412.80$, ant behavior bias: $\chi^2 = 316.36$, honey
252 bee behavior bias: $\chi^2 = 877.43$; $P < 0.001$ for all cases; N = 10520 in ant, N = 10011 in honey
253 bees). Genes that were loosely connected in co-expression networks constructed across all
254 samples tended to exhibit more caste and behavior bias in comparison to highly connected genes
255 (Fig. 4 *C* and *D*; Fig. S9 *C* and *D*). Genes with low connectivity are peripheral network elements,
256 while genes with high connectivity are central to regular networks and often highly pleiotropic
257 (59). Similarly, genes with high tissue-specificity across 12 honey bee tissues tended to exhibit
258 higher values caste and behavior bias in comparison to more pleiotropic, ubiquitously expressed
259 genes (Fig. S10), where tissue specificity was calculated using available data (46). Finally, genes
260 that were rapidly evolving (i.e. with high values of dN/dS) tended to exhibit higher levels of
261 caste and behavior bias, (Fig. 4 *E* and *F*; Fig. S9 *E* and *F*), with dN/dS estimated between each
262 focal species (*A. mellifera* and *M. pharaonis*) and a closely-related outgroup (*A. cerana* and *S.*
263 *invicta*, respectively).

264 We also detected strong negative correlations between gene-wise expression and overall
265 caste or behavior bias, (correlation of expression and ant caste bias: rho = -0.640; expression and
266 honey bee caste bias: rho = -0.676; expression and ant behavior bias: rho = -0.504; expression
267 and honey bee behavior bias: rho = -0.569; $P < 0.001$ for all; N = 10520 in ants , N = 10011 in
268 honey bees), where expression was averaged across the relevant comparisons, analogous to our
269 calculation of overall caste or behavior bias. However, our results are not solely driven by
270 expression levels. We performed partial correlation analysis accounting for expression to test for

271 the conditional effect of evolutionary age, connectivity, tissue specificity, and evolutionary rate
272 on caste or behavior bias. The relationship between caste/behavior bias and evolutionary age,
273 evolutionary rate, connectivity, and tissue-specificity generally all remained significant and in
274 the same direction when expression was accounted for (Table S11). The two exceptions were the
275 relationship between ant behavior bias and evolutionary age and honey bee caste bias and
276 connectivity (Table S11). However, when we removed the abdominal contribution to caste bias
277 and expression, the relationship between honey bee connectivity and caste bias remained
278 significant and negative (Table S11), revealing that the relationship between connectivity and
279 caste bias was universal except in the case of highly connected abdominally caste-biased genes,
280 as were present in our gene co-expression analysis (Fig. 2*B*). Together, these results indicate that
281 each gene's likelihood of recruitment to regulatory networks underlying caste or worker division
282 of labor is positively influenced by expression plasticity and position in overall regulatory
283 networks as well as the rate of coding sequence evolution and how recently such genes evolved.
284

285 **Discussion**

286 In this study we present the most comprehensive transcriptomic dataset to date of caste-
287 based division of labor for species from lineages representing two independent origins of
288 eusociality. We find generally low overlap between honey bees and pharaoh ants in the
289 thousands of genes that are differentially expressed between queen and worker larvae, pupae,
290 and adult tissues (Fig. 1*A*) as well as between nurses and foragers within the worker caste (Fig.
291 1*B*). This indicates that the independent evolution of these plastic phenotypes in honey bees and
292 ants has largely involved different genes. The one notable exception is in the abdomen, where we

293 find that about a third (~1500/4500) of all genes differentially expressed between queens and
294 workers in honey bees and ants have shared caste-biased expression in both species. This
295 proportion is nearly as high as the proportion of genes with shared expression bias across
296 development, which is considered to have deeply conserved molecular mechanisms (58, 65). We
297 find that genes that are upregulated in queens in both ants and honey bees also tend to be female-
298 biased in *D. melanogaster* (58), indicating that this set of shared caste-biased genes is derived
299 from ancient plastically-expressed genes underlying sexual dimorphism (Fig. 3*D*). Outside of
300 this large shared core set of ancient genes associated with female reproduction, the majority of
301 genes associated with caste and worker division of labor showed plastic expression in multiple
302 contexts (e.g., between developmental stages, tissues, and castes) and were relatively young,
303 rapidly evolving, and loosely connected in regulatory networks.

304 In the long-extinct solitary ancestors of social insects, female behaviors such as egg-
305 laying and brood care are thought to have been physiologically linked within an ovarian cycle
306 such that brood care followed egg-laying (22, 27). The reproductive groundplan hypothesis
307 proposes that these pathways existed in the common ancestor of eusocial species, and that caste-
308 based division of labor in eusocial lineages is derived from this reproductive groundplan (22, 24,
309 27, 28). However, past studies have failed to find large sets of genes repeatedly used for eusocial
310 evolution that would represent these pathways, and these studies have emphasized overlap in the
311 inferred functional categories (e.g., metabolism) of genes (37, 38, 40–42). Importantly, no
312 previous comparative study investigated caste-biased expression in the abdomen, the location of
313 the reproductive machinery where putative signatures of the reproductive groundplan are
314 expected to be most apparent.

315 In strong support of the reproductive groundplan hypothesis, we identify a large set
316 (~1500) of genes with shared caste-biased abdominal expression, many of which have known
317 reproductive or maternal function, including the receptor for *vitellogenin* as well as *vasa, smaug,*
318 and *ovo* (Fig. 2 *A* and *B*, Tables S8 and S9). This large overlap for abdominal caste-associated
319 genes is notable because honey bees and ants last shared a common ancestor approximately 160
320 million years ago (57), and this overlap is nearly as much as we see for genes that were
321 differentially expressed across developmental stages. While shared developmental molecular
322 mechanisms are presumably due simply to shared ancestry and the deep conservation of
323 developmental mechanisms (58, 65), the similar level of overlap for caste-associated genes
324 points to the large-scale recruitment of pre-existing developmental and physiological machinery
325 underlying the ancestral reproductive groundplan during the independent evolution of caste-
326 based division of labor in ant and honey bee lineages.
327 While reproductive caste in complex eusocial societies such as ants and honey bees is
328 typically fixed in adulthood, the tasks performed by workers (specifically, nursing versus
329 foraging) change over the course of the worker's adult lifetime (29, 49). This plastic behavioral
330 change is known to be accompanied by a wide range of physiological changes and is regulated at
331 least in part by conserved physiological pathways, for example, those involving insulin
332 signaling, juvenile hormone, and vitellogenin (29, 36, 66). However, we identified few genes
333 that were commonly differentially expressed between nurses and foragers in honey bees and
334 pharaoh ants (Fig. 1*B*), and the proportion of shared genes was much lower in comparison to
335 genes underlying abdominal differences between queens and workers. Nonetheless, we identified
336 a number of enriched Gene Ontology categories associated with development as well as

337 metabolism in each species (Tables S6 and S7), which is consistent with the notion that the
338 transition from nurse to forager is essentially a developmental process, and that common
339 molecular pathways may provide the raw genetic material for social evolution (25, 38, 42, 67).
340 While conserved factors or pathways clearly play important roles in aspects of caste
341 development and function as well as the transition from nursing to foraging, our results and other
342 studies indicate that the majority of the full transcriptomic architecture associated with caste and
343 age polyethism is not shared between species (38, 45, 49–51). Part of this lineage-specific
344 architecture is derived from expression differences in conserved genes: about half the
345 differentially expressed genes we identified had an ortholog in both species but did not exhibit
346 shared expression bias (Fig. 1 *A* and *B*), and the degree of caste-biased expression at particular
347 tissues or stages was generally only weakly correlated among orthologs of honey bees and ants
348 (Fig. S1). However, about 40% of caste-associated genes lacked 1:1 orthologs (either due to
349 apparent duplications or complete lack of orthology), reflecting the importance of taxonomically-
350 restricted genes for the evolution of eusociality (19, 46, 47, 68, 69). Additionally, we found that
351 orthologs with shared abdominal worker bias between honey bees and pharaoh ants were on
352 average evolutionarily younger (i.e. more taxonomically-restricted) than those with shared queen
353 bias (Fig. 1*C*), in line with previous studies (43–45, 48). Altogether, the emerging result of
354 sociogenomics research is that the bulk of the genetic architecture underlying caste-based
355 division labor is made up of plastically expressed genes with non-conserved sequence or
356 expression pattern (55, 64, 70, 71).
357 Novel polyphenisms are often thought to evolve from pre-existing developmental
358 plasticity, essentially by the canalization of previously continuous plastic phenotypes (21, 55,

359 72). Furthermore, morph-associated expression plasticity may arise largely neutrally (70), as
360 expression plasticity between morphs is often correlated to expression plasticity within morphs
361 (64, 73–76). Our results indicate that expression plasticity, both in highly conserved and
362 taxonomically-restricted genes, is a key feature of the genetic architecture underlying caste and
363 worker division of labor. Overall caste bias across 1:1 orthologs was correlated between *A.*
364 *mellifera* and *M. pharaonis*, where overall caste bias represents the likelihood of a gene to
365 exhibit expression bias in any comparison between queens and workers (Fig. S7*A*). Similarly,
366 overall behavior bias (overall bias between nurses and foragers) was correlated between species
367 (Fig. S7*B*), caste and behavior bias were themselves correlated (Fig. S8), and caste and behavior
368 bias were correlated to tissue specificity (Fig. S10). These results indicate that genes with plastic
369 expression in multiple contexts (across tissues, development, behavior, and caste) tend to be
370 consistently associated with caste and behavior in both honey bees and ants.

371 Genes with morph-specific effects on fitness are expected to evolve relatively rapidly due
372 to relaxed selection (77, 78), a prediction that has been supported in a number of organisms (64,
373 79–83). Elevated evolutionary rates may also pre-date the evolution of specific morphs (71, 74),
374 in support of the role of ancestral plasticity for the evolution of polyphenisms (21). Furthermore,
375 genes that are peripheral elements of regulatory networks and exhibit tissue-specific expression
376 tend to evolve rapidly due to relaxed selective constraint because they are less pleiotropic than
377 centrally located, ubiquitously expressed genes (75, 84–86). Finally, across organisms genes
378 identified as taxonomically-restricted tend to evolve more rapidly than older genes (87–89), but
379 rapidly evolving genes can mistakenly be identified as taxonomically-restricted genes by
380 BLASTP due to the rapid divergence of their amino acid sequence (90). Our results support all

381 of these empirical relationships, and we argue that they represent a set of characteristics of genes
382 which may not be readily disentangled. Evolutionary developmental biology emphasizes the
383 action of a certain type of gene for the evolution of morphological novelty: highly connected
384 core elements of that are highly pleiotropic and slowly evolving (65, 91). However, the emerging
385 picture is that the evolution of other traits such as social innovation also involves a majority of
386 genes on the opposite end of the spectrum: genes with highly variable expression across contexts
387 that are peripheral elements of regulatory networks and exhibit high levels of divergence in
388 expression and sequence between species (46, 49).
389
390 **Conclusions**
391 Our study shows that the recruitment of a large core of conserved reproductive-associated
392 genes, which make up a reproductive groundplan, is fundamental to the convergent evolution of
393 caste-based division of labor in ants and honey bees. However, our study also reveals that the
394 bulk of the full genetic architecture underlying the expression of social insect caste-based
395 division of labor varies between lineages. This is reflected by the general biology of social
396 insects, in that independently evolved societies share reproductive division of labor, the main
397 defining feature of eusociality, but also display a wide diversity of lineage-specific adaptations
398 (16). It is likely that a relatively small number of core conserved genes exist as upstream hubs in
399 regulatory networks, and layered on top of this core is a myriad of taxonomically-restricted
400 genes as well as conserved genes with lineage-specific expression patterns (46, 49, 92, 93). This
401 is consistent with models for the evolution of hierarchical developmental gene regulatory
402 networks, whereby a relatively small number of highly conserved genes act upstream to initiate

403 gene cascades (e.g., to set up body-patterning), while batteries of downstream genes are much
404 more evolutionarily labile, and are largely responsible for lineage-specific features (94, 95).
405 Recent studies have made progress elucidating the function of several core candidate genes and
406 gene pathways that regulate caste-based division of labor by manipulating individual genes (34–
407 36, 96). A complementary goal of such candidate gene-based research is to use unbiased
408 approaches to identify the full suite of genes underlying caste-based division of labor in multiple
409 independent lineages and subsequently to ask what proportion of genes show conserved function
410 across lineages. Large-scale transcriptomic studies such as ours are an invaluable starting point
411 to achieve this goal. However, transcriptomic studies can only hint at the relative functional
412 importance of genes, and future studies will have to assess the functional importance of various
413 components of the transcriptomic architecture, including the relative importance of of shared
414 versus lineage-specific genetic mechanisms for the convergent evolution of eusociality.
415

416 **Methods**

417 *Study Design*

418 We collected parallel time series RNA-seq data of caste development in the pharaoh ant
419 *Monomorium pharaonis* and the honey bee *Apis mellifera*, including seven developmental stages
420 (egg, five larval stages, one pupal stage) plus each of three adult body segments in both species
421 (see Table S1 for list of all samples). We separated adults into the three main body segments
422 (head, mesosoma, and metosoma) upon sample collection and sequenced pools of each part
423 separately. For convenience, we refer to these segments as “head”, “thorax”, and “abdominal”
424 tissues throughout. We sequenced whole embryos and whole bodies of larvae and pupae. Each

425 sample represents a pool of ten individuals taken from the same colony, and replicate samples of
426 the same type represent pools of individuals taken from different colonies.
427
428 *Ant collection*
429 To collect samples of *M. pharaonis* across development, we created 27 replicate colonies of
430 ~400 workers and ~400 total larvae from a large mixed genetic source. We removed queens from
431 each colony, which stimulates the production of new queens and males from existing eggs and
432 L1 larvae in *M. pharaonis* (97–99). We pre-assigned each colony to one of nine sample types,
433 ordered by developmental timing (egg, L1-L5 larvae, pupae, virgin queens/males,
434 nurses/foragers). We allowed the 27 colonies to grow for four weeks, and collected samples
435 progressively when the youngest individuals left in the colonies represented the assigned
436 developmental stage (note that *M. pharaonis* workers, lacking ovaries (100), do not begin to lay
437 eggs, so the brood progressively ages as no replacement eggs are laid).
438 We identified larval stage and caste as in (101), by hair and morphology (102). To
439 synchronize pupal developmental collection, we exclusively sampled pupae whose eyes had
440 darkened. We sampled males as soon as they had eclosed as adults from the pupal stage. We
441 distinguished between nurses and foragers based on behavioral observation. Specifically, we
442 observed nurses feeding larvae and we observed foragers collecting food. We sampled egg-
443 laying mature queens (3-4 months old) from the group of queens we initially removed to
444 stimulate reproduction to ensure that queens and workers came from the same genetic
445 background.
446

447 *Honey bee collection*

448 To collect samples of *A. mellifera* across development, we established experimental colonies in
449 which queens were allowed to lay directly onto empty comb for 24 hours to ensure control of
450 larval age and that larvae of a given replicate were from the same queen. We collected eggs after
451 this period directly from the comb. After three days, we grafted a subset of hatched larvae into
452 artificial queen cells in queenless portions of the hive. Starting at that day, we sampled the five
453 stages of larvae (L1-L5) on each consecutive day. We sampled pupae once their eyes had
454 darkened to synchronize developmental timing. We sampled males by placing "hair-roller" cages
455 on top of cells during pupation and waiting for individuals to emerge from pupation. We sampled
456 egg-laying mature queens from separate colonies and pooled them into replicate samples. We
457 paint-marked worker individuals upon emergence from the pupal stage and sampled nurses that
458 were less than 7 days old and foragers that were greater than 21 days old.

459

460 *RNA extraction, sequencing, aligning to genomes*

461 We isolated RNA using Trizol reagents. We performed cDNA synthesis and library preparation
462 were performed using a previously described protocol (Aird et al. 2017), with the only alteration
463 being that the input RNA was 50ng and the cycle number of cDNA amplification was increased
464 to sixteen. To compare sample quality across the experiment and test our ability to detect lowly-
465 expressed genes, we added ERCC92 (Thermo Fisher Scientific Inc.) spike-in mixes to total RNA
466 prior to amplification. We pooled libraries with an equal amount of cDNA and sequenced single-
467 end for 50 cycles in Illumina Hiseq 2500. We aligned reads to reference genomes using Bowtie2
468 (103). *A. mellifera* reads were aligned to NCBI gene models, genome version 4.5, and *M.*

469 *pharaonis* reads were aligned to NCBI gene models, genome version 2.0. We estimated read
470 count and transcripts per million (TPM) using RSEM (104).

471

472 *Identification of orthologs*

473 To identify orthologs between *A. mellifera* and *M. pharaonis*, we started with a curated
474 orthology map of *Aculeata* species from OrthoDB9 (Zdobnov et al 2016). We downloaded
475 amino acid sequences for each species from RefSeq (105). We associated transcripts with
476 OrthoDB9 protein names using BLASTp (E-value 10^{-10}) and identified the Aculeata ortholog
477 group matched by each gene based on the identified BLASTp hits. In this way, we identified 1:1,
478 one-to-many, and many-to-many orthologous groups between *A. mellifera* and *M. pharaonis*. For
479 direct comparison of the species, we restricted our analysis to 1:1 orthologs (i.e. genes for which
480 only one gene from each species matches the given OrthoDB9 ortholog group). We identified
481 three-way 1:1:1 orthologs between *A. mellifera*, *M. pharaonis*, and *Drosophila melanogaster*
482 using a similar procedure based on Endopterygota orthology groups from OrthoDB9.

483

484 *Differential Expression Analysis*

485 To identify caste-associated differentially-expressed genes (DEGs), we performed differential
486 expression analysis between queens and workers at each developmental stage and tissue,
487 separately for each species. We constructed GLM-like models including replicate and caste and
488 identified genes associated with caste (FDR < 0.1) at each stage or tissue using EdgeR (106).
489 Similarly, to identify behavioral DEGs we performed differential expression analysis between

490 nurses and foragers for each tissue. To identify developmental DEGs in each species, we
491 constructed models with all larval and egg samples and identified genes differentially expressed
492 between any developmental stage (FDR < 0.1), controlling for overall caste differences.
493
494 *Gene co-expression analysis*
495 In contrast to many network methods which assess gene-gene relationships across all samples,
496 biclustering seeks to identify a group of genes which are coexpressed (i.e. exhibit concerted
497 expression changes) across a subset of sample types (107). Given that our data contained a large
498 number of sample types, we reasoned that we could employ biclustering to identify groups of
499 genes particularly associated with a given sample type. We performed plaid clustering, one of the
500 top performing biclustering algorithms in a recent survey (108). We used the R package biclust
501 to implement clustering (109). Plaid clustering models expression level for each gene as a
502 function of bicluster weights, where only biclusters containing the gene contribute to predicted
503 expression level (110, 111). The algorithm iteratively constructs layers containing samples and
504 genes and retains layers that improve the model fit, where layers represent biclusters.
505 Plaid clustering is non-deterministic and individual biclusters are not found in every
506 iteration of clustering. To define a reasonable ensemble of biclusters, we performed clustering
507 1000 times separately for each species, using inverse hyperbolic sine transformed TPM
508 (transcripts per million) (112). While a large number of interesting bicluster definitions are
509 possible, we decided to identify biclusters that consistently contained all queen abdomen samples
510 to focus our investigation on the tissue that exhibited the strongest signature of caste bias.
511 Specifically, we extracted biclusters containing all three mature queen abdomen samples and no

512 more than three other samples total (note that honey bee queen abdomen samples clustered with
513 egg samples, while pharaoh ant queen samples did not cluster with egg samples).
514 Because the same genes were not always present in such a bicluster, we tabulated the
515 number of queen abdomen biclusters each gene was found in and retained genes present in a
516 higher proportion of biclusters than a given cut-off, determined by inspection of frequency
517 distributions of bicluster presence. In pharaoh ants, we found a large set of genes present in
518 greater than 90% of queen abdomen biclusters, and we retained these genes for further analysis
519 (N = 1039 genes; Fig. S11A, i.e. the same set of genes was repeatedly found). In contrast, honey
520 bee queen abdomen biclusters tended to contain one of two groups of genes, as the frequency of
521 presence in the bicluster peaks at 60% and 30% (Fig. S11B). Out of 1245 genes present in
522 greater than 60% of the identified biclusters, 877 were differentially expressed and upregulated
523 in queen abdomens relative to worker (also note that this set of genes exhibited much higher
524 expression in eggs than the latter set). In contrast, out of 1057 genes present in 25-35% of
525 biclusters, 611 out of were differentially expressed and upregulated in worker abdomens,
526 compared to 47 upregulated in queen abdomens. Therefore, it is clear that the more common
527 bicluster represents genes associated with queen abdomens, so we retained this set of genes for
528 further analysis (N = 1245 genes).
529 We proceeded with our analysis using these identified sets of genes, which we term
530 modules associated with queen abdominal expression. We calculated connectivity in the module
531 (i.e. intra-module connectivity) as the sum of pairwise Pearson correlations, where correlation
532 values are raised to the sixth power, the standard value for unsigned weighted gene co-
533 expression networks (113). A major goal of gene co-expression analysis is the identification of

534 hub genes, genes central to networks that are strongly associated to relevant traits (59). To this
535 end, we conservatively identified hub genes associated with queen abdominal expression as
536 genes with intra-module connectivity in at least the 90th percentile and abdominal $\log_2$ fold-
537 change values greater than 2 (representing a 4-fold increase in expression in queen relative to
538 worker abdomens).
539
540 *Phylostratigraphy*
541 We estimated the evolutionary age of each gene using phylostratigraphy. Phylostratigraphy
542 groups genes into hierarchical age categories based on identifiable orthology (using BLASTp)
543 (114, 115). For example, genes found in ants and honey bees but not in non-aculeate
544 hymenopterans would be labeled "Aculeata" genes, while genes shared between vertebrates and
545 insects would be labeled "Bilateria". For our purposes, we decided to focus on the difference
546 between "ancient" genes, which we defined as displaying orthology with non-insect animals, and
547 a number of hierarchical younger categories: "insect", "hymenopteran", "aculeate", "ant", "bee",
548 and "novel" (where "ant" refers to genes found in *M. pharaonis* and other ants but not in any
549 other species, "bee" refers to genes found in *A. mellifera* and other bees but not in other species,
550 and "novel" refers to a gene found only in *A. mellifera* or *M. pharaonis*).
551 A key component of phylostratigraphy is the creation of a BLAST database in which to
552 identify orthologs (114, 115). Because we largely planned to focus on younger age categories,
553 we constructed a protein database containing all annotated hymenopteran genomes (48 total). We
554 added to this group ten non-hymenopteran insect genomes and ten non-arthropod genomes (see
555 Table S12 for a full list of included genomes). Therefore, a gene labeled as "ancient" displayed a

556 significant BLASTp hit to one of the ten non-arthropod genomes. While phylostratigraphy
557 typically employs an extremely large database containing all available representative taxa, we
558 reasoned that for our study resolution between categories such as "Bilateria" and "Eukaryota"
559 was unnecessary. Furthermore, adding extraneous genomes effectively dilutes the database, such
560 that more similarity is needed to pass an E-value threshold. Because we included only a sample
561 of non-hymenopteran genomes, we were therefore able to stringently identify orthologs (E-value
562 10^{-10} in comparison to a typical value of 10^{-5} (116)) and accurately place them along the
563 hymenopteran phylogeny.
564
565 *Comparison to Drosophila melanogaster*
566 We downloaded available MODENCODE RNA-seq data on male and female whole bodies of *D.*
567 *melanogaster* (58). Male and female samples came from 5- and 30-day old flies (one replicate
568 for each sex). While much recent work has detailed tissue-specific (and even cell-specific)
569 expression in *D. melanogaster*, we reasoned that whole bodies (lacking the same body parts as
570 we used in this study) would give us reasonable overall levels of sex bias to compare our data to.
571 We aligned reads to the current release of the *D. melanogaster* genome available on NCBI
572 (*assembly release 6 plus ISO*) using the same pipeline as described for *A. mellifera* and *M.*
573 *pharaonis*, and we performed differential expression analysis between males and females as
574 described above.
575
576 *Calculation of overall bias*

577 Previous work estimated the overall degree to which genes exhibited morph bias in RNA-seq
578 data among four morphs by calculating the Euclidean distance of $\log_2$ fold-change between each
579 comparison (64). Following this reasoning, we calculated overall caste bias as the Euclidean
580 distance of $\log_2$ fold-change between queens and workers at each larval developmental stage
581 (L2-L5), pupae, and each adult tissue. Similarly, we calculated overall behavior bias as the
582 Euclidean distance of $\log_2$ fold-change between nurses and foragers in each tissue. To perform
583 partial correlations controlling for magnitude of gene expression with overall bias as the
584 dependent variable, we calculated expression analogously, as the Euclidean distance of $\log_{10}$
585 counts-per-million at each stage/tissue tested.
586
587 *Estimation of tissue specificity*
588 While comparing expression between castes across the entirety of development in many tissues
589 was unfeasible for this study, a previous study performed RNA-sequencing on twelve tissues in
590 *A. mellifera* in worker nurses and foragers (46). We downloaded the available data and aligned
591 reads to the current version of the *A. mellifera* genome, using the pipeline described above. To
592 classify genes by their tissue specificity, we calculated τ, a commonly used metric of expression
593 specificity (117). τ ranges from 0 to 1, where 0 indicates genes are ubiquitously expressed and 1
594 indicates genes are exclusively expressed in one tissue.
595
596 *Estimation of dN/dS*
597 We estimated evolutionary rate using dN/dS, the ratio of non-synonymous to synonymous
598 nucleotide changes. We estimated pairwise dN/dS between each focal species and a second

599 closely related species. For *A. mellifera*, we compared protein-coding sequences to *A. cerana*.
600 For *M. pharaonis*, we compared protein-coding sequences to *Solenopsis invicta*. We identified
601 1:1 orthologs using OrthoDB9, as described above. For each 1:1 ortholog pair, we selected the
602 longest transcript associated with the gene for each pair of species. We aligned orthologous
603 protein sequences using ClustalW (118), derived nucleotide alignments from protein alignments
604 using pal2nal (119), and estimated pairwise dN/dS of nucleotide alignments using PAML,
605 package codeml (120).
606
607 *Gene Ontology Analysis*
608 We performed Gene Set Enrichment Analysis (GSEA) using the R package topGO (121). Rather
609 than relying on automatically annotated Gene Ontology (GO) terms, we utilized the well-curated
610 *D. melanogaster* gene ontology database, downloaded from FlyBase (122). We performed GSEA
611 analysis on genes with 1:1 orthologs identified previously, associating the *D. melanogaster* GO
612 terms to *A. mellifera* and *M. pharaonis* orthologs. We identified GO terms associated with caste-
613 biased or behavior-biased genes using the P-value of differential expression between queens and
614 workers or nurses and foragers. We identified GO terms associated with overall caste or behavior
615 bias using the value of overall caste bias or overall behavior bias as input. We identified enriched
616 terms with P-value < 0.05.
617
618 *General statistical analyses*
619 We performed all statistical analyses and constructed plots in R, version 3.4.0 (R Core Team
620 2017), aided by the packages “plyr” (123), “ggplot2” (124), “gridExtra”, version 2.3 (125) and

621 “cowplot”, version 0.9.3 (126). To test for the effect of phylostrata on caste or behavior bias, we
622 constructed a generalized linear model using the Gamma distribution and treating phylostrata as
623 an ordinal variable. To test for the effects of connectivity and evolutionary rate (dN/dS) on caste
624 or behavior bias while controlling for expression, we employed partial correlation analysis using
625 the package “ppcor” (127).
626
627 *Data availability*
628 All data and scripts required to generate figures, tables, and perform statistical analyses are
629 available on Github: https://github.com/oist/devnetwork. Raw reads are deposited at DNA Data
630 Bank of Japan.
631
632 **Acknowledgements**
633 We thank Luigi Pontieri for photographs of pharaoh ants, and Alex Wild for photographs of
634 honey bees. We thank Chao Tong for compiling hymenopteran genomes for use in
635 phylostratigraphy and for comments on the manuscript. We thank Rohini Singh for comments on
636 the manuscript. We thank Junhyong Kim, Mia Levine, and Justin Walsh for helpful comments
637 and discussion. This work was funded by the National Science Foundation (grant number IOS-
638 1452520 to TAL) and United States Department of Agriculture National Institute of Food and
639 Agriculture (grant number 2014-67013-21725 to TAL), and subsidy funding from Okinawa
640 Institute of Technology to ASM. The National Science Foundation also funded MRW (DGE-
641 1321851).
642

643 **References**

644 1. Stern DL (2013) The genetic causes of convergent evolution. *Nat Rev Genet* 14(11):751–
645 764.

646 2. Stern DL, Orgogozo V (2008) The loci of evolution: how predictable is genetic evolution?
647 *Evolution* 62(9):2155–2177.

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

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

652 5. Parker J, et al. (2013) Genome-wide signatures of convergent evolution in echolocating
653 mammals. *Nature* 502(7470):228–231.

654 6. Foote AD, et al. (2015) Convergent evolution of the genomes of marine mammals. *Nat*
655 *Genet* 47(3):272–275.

656 7. Thomas GWC, Hahn MW (2015) Determining the null model for detecting adaptive
657 convergence from genomic data: a case study using echolocating mammals. *Mol Biol Evol*
658 32(5):1232–1236.

659 8. Zou Z, Zhang J (2015) No genome-wide protein sequence convergence for echolocation.
660 *Mol Biol Evol* 32(5):1237–1241.

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

663 10. Scharff C, Petri J (2011) Evo-devo, deep homology and *FoxP2*: implications for the
664 evolution of speech and language. *Philos Trans R Soc Lond B Biol Sci* 366(1574):2124–
665 2140.

666 11. Halder G, Callaerts P, Gehring WJ (1995) Induction of ectopic eyes by targeted expression
667 of the eyeless gene in *Drosophila*. *Science* 267(5205):1788–1792.

668 12. Gehring WJ (2005) New perspectives on eye development and the evolution of eyes and
669 photoreceptors. *J Hered* 96(3):171–184.

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

672 14. Khalturin K, et al. (2008) A novel gene family controls species-specific morphological traits
673 in Hydra. *PLoS Biol* 6(11):e278.

674 15. Babonis LS, Martindale MQ, Ryan JF (2016) Do novel genes drive morphological novelty?
675 An investigation of the nematosomes in the sea anemone *Nematostella vectensis*. *BMC Evol*
676 *Biol* 16(1):114.

677 16. Wilson EO (1971) *The insect societies*. (Harvard University Press).

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

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

682 19. Johnson BR, Linksvayer TA (2010) Deconstructing the superorganism: social physiology,
683 groundplans, and sociogenomics. *Q Rev Biol* 85(1):57–79.

684 20. Hölldobler B, Wilson EO (2009) *The Superorganism: The Beauty, Elegance, and*
685 *Strangeness of Insect Societies* (W. W. Norton & Company).

686 21. Nijhout HF (2003) Development and evolution of adaptive polyphenisms. *Evol Dev* 5(1):9–
687 18.

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

690 23. Linksvayer TA, Wade MJ (2005) The evolutionary origin and elaboration of sociality in the
691 aculeate Hymenoptera: maternal effects, sib-social effects, and heterochrony. *Q Rev Biol*
692 80(3):317–336.

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

695 25. Toth AL, Robinson GE (2007) Evo-devo and the evolution of social behavior. *Trends Genet*
696 23(7):334–341.

697 26. Toth AL, et al. (2010) Brain transcriptomic analysis in paper wasps identifies genes
698 associated with behaviour across social insect lineages. *Proc Biol Sci* 277(1691):2139–
699 2148.

700 27. West-Eberhard MJ (1996) *Natural history and evolution of paper-wasps* (Oxford University
701 Press).

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

705 29. Robinson GE (1987) Regulation of honey bee age polyethism by juvenile hormone. *Behav*

706 *Ecol Sociobiol* 20(5):329–338.

707 30. Ament SA, Corona M, Pollock HS, Robinson GE (2008) Insulin signaling is involved in the
708 regulation of worker division of labor in honey bee colonies. *Proc Natl Acad Sci U S A*
709 105(11):4226–4231.

710 31. Libbrecht R, et al. (2013) Interplay between insulin signaling, juvenile hormone, and
711 vitellogenin regulates maternal effects on polyphenism in ants. *Proc Natl Acad Sci U S A*
712 110(27):11050–11055.

713 32. Wheeler DE, Buck N, Evans JD (2006) Expression of insulin pathway genes during the
714 period of caste determination in the honey bee, *Apis mellifera*. *Insect Mol Biol* 15(5):597–
715 602.

716 33. Okada Y, Watanabe Y, Tin MMY, Tsuji K, Mikheyev AS (2017) Social dominance alters
717 nutrition-related gene expression immediately: transcriptomic evidence from a
718 monomorphic queenless ant. *Mol Ecol* 26(11):2922–2938.

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

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

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

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

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

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

733 40. Woodard SH, Bloch GM, Band MR, Robinson GE (2014) Molecular heterochrony and the
734 evolution of sociality in bumblebees (*Bombus terrestris*). *Proc Biol Sci*
735 281(1780):20132419.

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

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

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

743 44. Warner MR, Mikheyev AS, Linksvayer TA (2017) Genomic signature of kin selection in an
744 ant with obligately sterile workers. *Mol Biol Evol* 34(7):1780–1787.

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

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

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

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

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

756 50. Morandin C, et al. (2016) Comparative transcriptomics reveals the conserved building
757 blocks involved in parallel evolution of diverse phenotypic traits in ants. *Genome Biol*
758 17:43.

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

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

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

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

767 55. Jones BM, Kingwell CJ, Wcislo WT, Robinson GE (2017) Caste-biased gene expression in
768 a facultatively eusocial bee suggests a role for genetic accommodation in the evolution of
769 eusociality. *Proc Biol Sci* 284(1846).

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

772 57. Peters RS, et al. (2017) Evolutionary history of the Hymenoptera. *Curr Biol* 27(7):1013–
773 1018.

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

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

778 60. Barchuk AR, Bitondi MMG, Simões ZLP (2002) Effects of juvenile hormone and ecdysone
779 on the timing of vitellogenin appearance in hemolymph of queen and worker pupae of *Apis*
780 *mellifera*. *J Insect Sci* 2:1.

781 61. Khila A, Abouheif E (2010) Evaluating the role of reproductive constraints in ant social
782 evolution. *Philos Trans R Soc Lond B Biol Sci* 365(1540):617–630.

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

785 63. Mével-Ninio M, Terracol R, Salles C, Vincent A, Payre F (1995) *ovo*, a *Drosophila* gene
786 required for ovarian development, is specifically expressed in the germline and shares most
787 of its coding sequences with *shavenbaby*, a gene involved in embryo patterning. *Mech Dev*
788 49(1):83–95.

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

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

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

795 67. Rittschof CC, Robinson GE (2016) Behavioral genetic toolkits: toward the evolutionary
796 origins of complex phenotypes. *Curr Top Dev Biol* 119:157–204.

797 68. Sumner S (2014) The importance of genomic novelty in social evolution. *Mol Ecol*
798 23(1):26–28.

799 69. Simola DF, et al. (2013) Social insect genomes exhibit dramatic evolution in gene
800 composition and regulation while preserving regulatory features linked to sociality. *Genome*
801 *Res* 23(8):1235–1247.

802 70. Helanterä H, Uller T (2014) Neutral and adaptive explanations for an association between
803 caste-biased gene expression and rate of sequence evolution. *Front Genet* 5:297.

804 71. Hunt BG, et al. (2011) Relaxed selection is a precursor to the evolution of phenotypic
805 plasticity. *Proc Natl Acad Sci U S A* 108(38):15936–15941.

806 72. Moczek AP, et al. (2011) The role of developmental plasticity in evolutionary innovation.
807 *Proc Biol Sci* 278(1719):2705–2713.

808 73. Hunt BG, Ometto L, Keller L, Goodisman MAD (2013) Evolution at two levels in fire ants:
809 the relationship between patterns of gene expression and protein sequence evolution. *Mol*
810 *Biol Evol* 30(2):263–271.

811 74. Leichty AR, Pfennig DW, Jones CD, Pfennig KS (2012) Relaxed genetic constraint is
812 ancestral to the evolution of phenotypic plasticity. *Integr Comp Biol* 52(1):16–30.

813 75. Mank JE, Ellegren H (2009) Are sex-biased genes more dispensable? *Biol Lett* 5(3):409–
814 412.

815 76. Mank JE, Hultin-Rosenberg L, Axelsson E, Ellegren H (2007) Rapid evolution of female-
816 biased, but not male-biased, genes expressed in the avian brain. *Mol Biol Evol* 24(12):2698–
817 2706.

818 77. Van Dyken JD, Wade MJ (2010) The genetic signature of conditional expression. *Genetics*
819 184(2):557–570.

820 78. Snell-Rood EC, Van Dyken JD, Cruickshank T, Wade MJ, Moczek AP (2010) Toward a
821 population genetic framework of developmental evolution: the costs, limits, and
822 consequences of phenotypic plasticity. *Bioessays* 32(1):71–81.

823 79. Demuth JP, Wade MJ (2007) Maternal expression increases the rate of *bicoid* evolution by
824 relaxing selective constraint. *Genetica* 129(1):37–43.

825 80. Cruickshank T, Wade MJ (2008) Microevolutionary support for a developmental hourglass:
826 gene expression patterns shape sequence variation and divergence in *Drosophila*. *Evol Dev*
827 10(5):583–590.

828 81. Purandare SR, Bickel RD, Jaquiery J, Rispe C, Brisson JA (2014) Accelerated evolution of
829 morph-biased genes in pea aphids. *Mol Biol Evol* 31(8):2073–2083.

830 82. Snell-Rood EC, et al. (2011) Developmental decoupling of alternative phenotypes: insights
831 from the transcriptomes of horn-polyphenic beetles. *Evolution* 65(1):231–245.

832 83. Pespeni MH, Ladner JT, Moczek AP (2017) Signals of selection in conditionally expressed
833 genes in the diversification of three horned beetle species. *J Evol Biol* 30(9):1644–1657.

834 84. Hahn MW, Kern AD (2005) Comparative genomics of centrality and essentiality in three
835 eukaryotic protein-interaction networks. *Mol Biol Evol* 22(4):803–806.

836 85. Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW (2002) Evolutionary rate in
837 the protein interaction network. *Science* 296(5568):750–752.

838 86. Liao B-Y, Scott NM, Zhang J (2006) Impacts of gene essentiality, expression pattern, and
839 gene compactness on the evolutionary rate of mammalian proteins. *Mol Biol Evol*
840 23(11):2072–2080.

841 87. Albà MM, Castresana J (2005) Inverse relationship between evolutionary rate and age of
842 mammalian genes. *Mol Biol Evol* 22(3):598–606.

843 88. Wolf YI, Novichkov PS, Karev GP, Koonin EV, Lipman DJ (2009) The universal
844 distribution of evolutionary rates of genes and distinct characteristics of eukaryotic genes of
845 different apparent ages. *Proc Natl Acad Sci U S A* 106(18):7273–7280.

846 89. Cai JJ, Woo PCY, Lau SKP, Smith DK, Yuen K-Y (2006) Accelerated evolutionary rate
847 may be responsible for the emergence of lineage-specific genes in ascomycota. *J Mol Evol*
848 63(1):1–11.

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

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

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

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

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

859 95. Peter IS, Davidson EH (2015) *Genomic control process*. (Academic Press).

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

862 97. Peacock AD, Baxter AT (1950) Studies in pharaoh's ant, *Monomorium pharaonis (L.)*, 3:
863 life history. *Entomol Mon Mag* 86:171–178.

864 98. Edwards JP (1987) Caste regulation in the pharaoh's ant *Monomorium pharaonis*: the
865 influence of queens on the production of new sexual forms. *Physiol Entomol* 12(1):31–39.

866 99. Warner MR, Lipponen J, Linksvayer TA (2018) Pharaoh ant colonies dynamically regulate
867 reproductive allocation based on colony demography. *Behav Ecol Sociobiol* 72(3):31.

868 100. Oster GF, Wilson EO (1978) Caste and ecology in the social insects. *Monogr Popul Biol*
869 12:1–352.

870 101. Warner MR, Kovaka K, Linksvayer TA (2016) Late-instar ant worker larvae play a
871 prominent role in colony-level caste regulation. *Insectes Soc* 63(4):575–583.

872 102. Berndt KP, Kremer G (1986) Larvenmorphologie der Pharaoameise *Monomorium*
873 *pharaonis (L.)* (Hymenoptera, Formicidae). *Zool Anz* 216.

874 103. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods*
875 9(4):357–359.

876 104. Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with
877 or without a reference genome. *BMC Bioinformatics* 12:323.

878 105. O'Leary NA, et al. (2016) Reference sequence (RefSeq) database at NCBI: current status,
879 taxonomic expansion, and functional annotation. *Nucleic Acids Res* 44(D1):D733–45.

880 106. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for
881 differential expression analysis of digital gene expression data. *Bioinformatics* 26(1):139–
882 140.

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

885 108. Oghabian A, Kilpinen S, Hautaniemi S, Czeizler E (2014) Biclustering methods: biological
886 relevance and application in gene expression analysis. *PLoS One* 9(3):e90801.

887 109. Kaiser S, Leisch F (2008) A toolbox for bicluster analysis in R.

888 110. Turner H, Bailey T, Krzanowski W (2005) Improved biclustering of microarray data
889 demonstrated through systematic performance tests. *Comput Stat Data Anal* 48(2):235–254.

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

891 112. Brawand D, et al. (2011) The evolution of gene expression levels in mammalian organs.
892 *Nature* 478(7369):343–348.

893 113. Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network
894 analysis. *BMC Bioinformatics* 9:559.

895 114. Domazet-Lošo T, Tautz D (2010) A phylogenetically based transcriptome age index mirrors
896 ontogenetic divergence patterns. *Nature* 468(7325):815–818.

897 115. Domazet-Lošo T, Brajković J, Tautz D (2007) A phylostratigraphy approach to uncover the
898 genomic history of major adaptations in metazoan lineages. *Trends Genet* 23(11):533–539.

899 116. Drost H-G, Gabel A, Grosse I, Quint M (2015) Evidence for active maintenance of
900 phylotranscriptomic hourglass patterns in animal and plant embryogenesis. *Mol Biol Evol*
901 32(5):1221–1231.

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

904 118. Larkin MA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23(21):2947–
905 2948.

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

909 120. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*
910 24(8):1586–1591.

911 121. Alexa A, Rahnenfuhrer J (2010) topGO: enrichment analysis for gene ontology.

912 122. Gramates LS, et al. (2016) FlyBase at 25: looking to the future. *Nucleic Acids*
913 *Res*:gkw1016.

914 123. Wickham H, Others (2011) The split-apply-combine strategy for data analysis. *J Stat Softw*
915 40(1):1–29.

916 124. Wickham H (2016) *ggplot2: Elegant Graphics for Data Analysis* (Springer).

917 125. Auguie B (2012) gridExtra: functions in Grid graphics. *R package version 0 9* 1.

918 126. Wilke CO (2016) cowplot: streamlined plot theme and plot annotations for "ggplot2".

919 127. Kim S (2015) ppcor: An R package for a fast calculation to semi-partial correlation
920 coefficients. *Commun Stat Appl Methods* 22(6):665–674.

921 128. Schmidt AM, Linksvayer TA, Boomsma JJ, Pedersen JS (2011) Queen–worker caste ratio
922 depends on colony size in the pharaoh ant (*Monomorium pharaonis*). *Insectes Soc*
923 58(2):139–144.

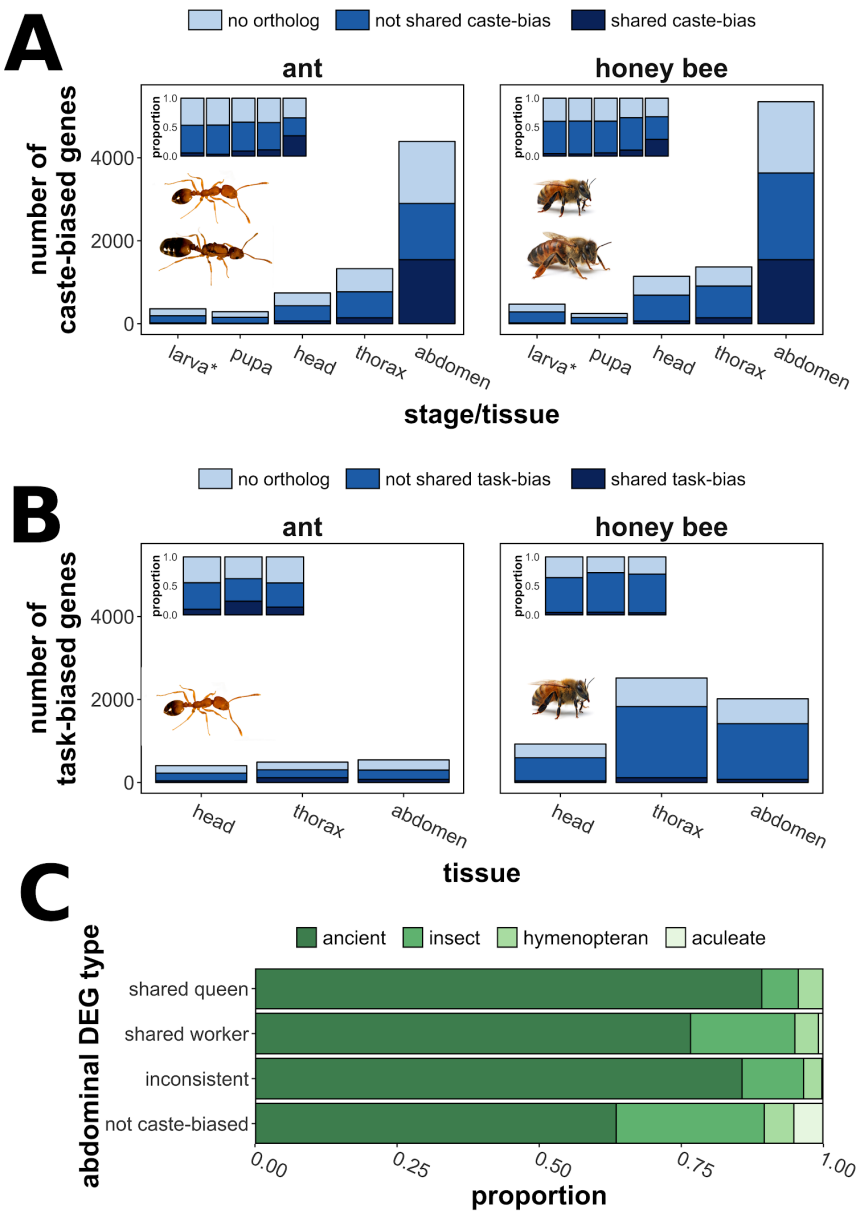

Fig. 1. Number of differentially expressed genes (FDR < 0.1) between A) queens and workers and B) nurses and foragers at each developmental stage or tissue in *M. pharaonis* (left) and *A. mellifera* (right). "Head", "thorax", and "abdomen" refer to body segments of adults, while pupa and larva refers to whole bodies. Genes are divided by color into categories based on gene-gene orthology between the two species. "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) Across 1:1 orthologs, genes with shared queen bias (upregulated in queen abdomens of both species) are comprised of proportionally more ancient genes than those with shared worker bias or genes with inconsistent bias.

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

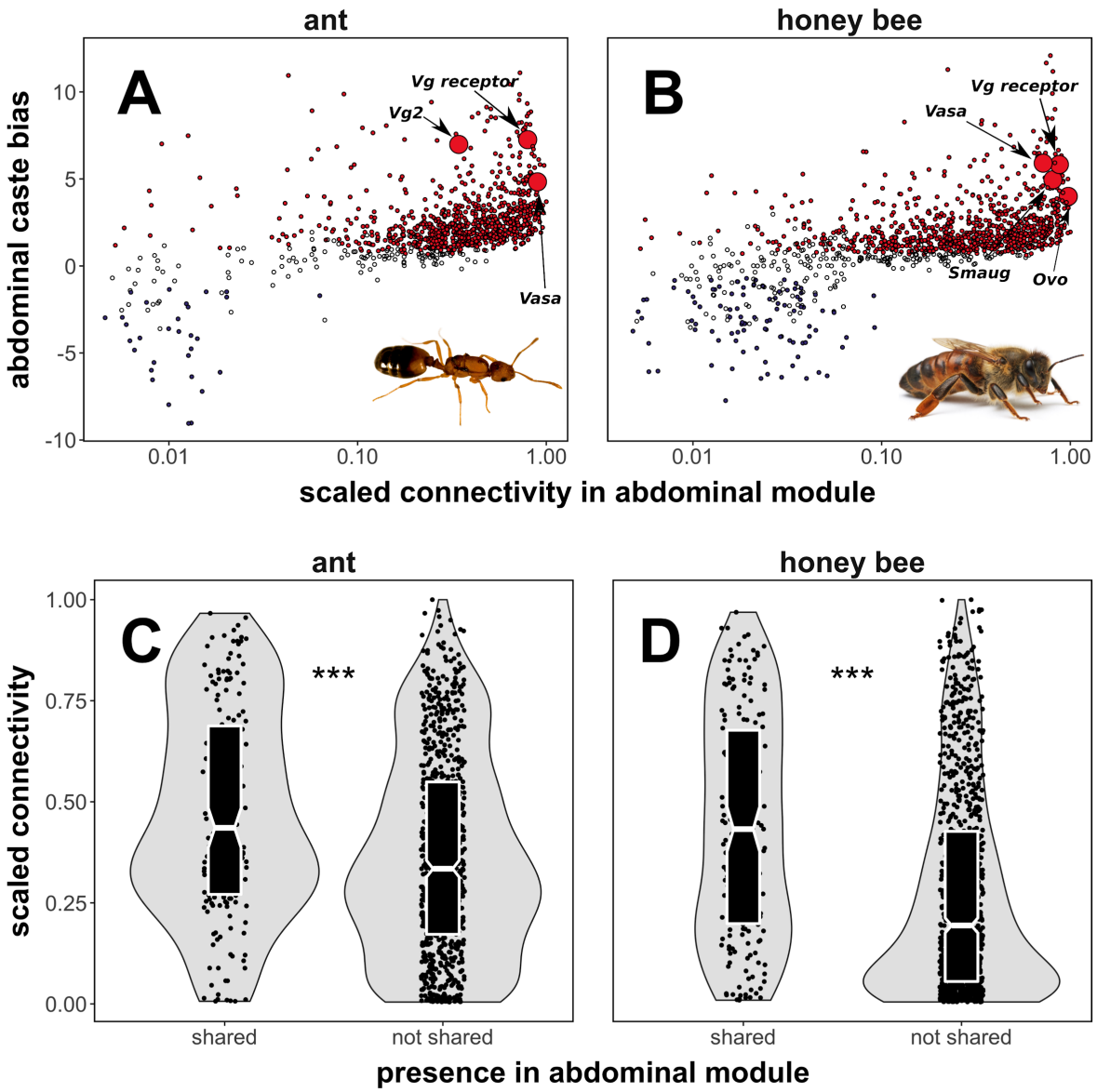

Fig. 2. Abdominal caste bias ($\log_2$ fold change queen versus worker) is correlated with connectivity within the queen-abdomen module in A) ants (rho = 0.536, P < 0.001) and B) honey bees (rho = 0.617, P < 0.001), indicating that centrally located genes in the module are highly queen-biased. Genes 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)

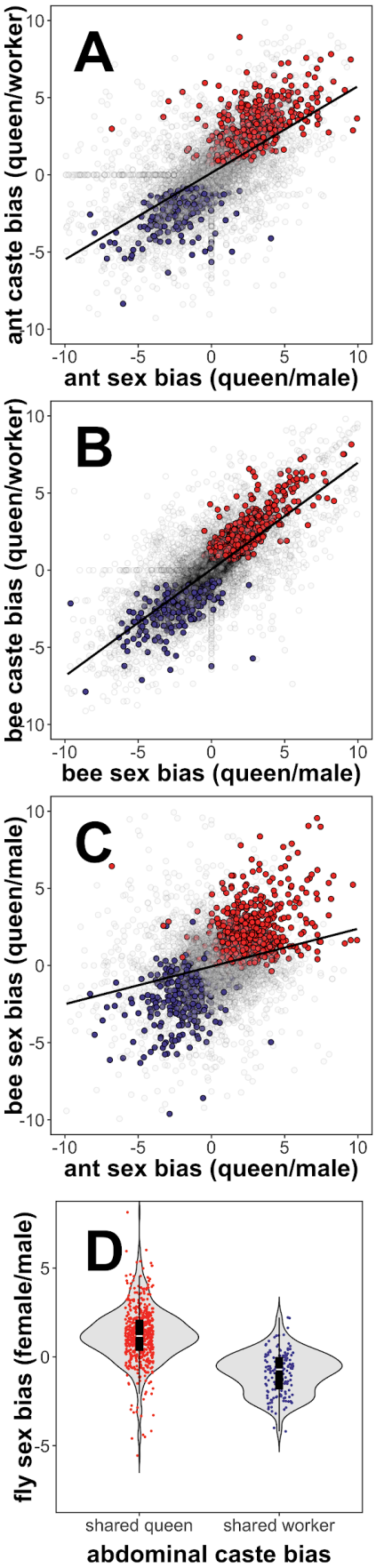
A
ant caste bias (queen/worker)
ant sex bias (queen/male)
B
bee caste bias (queen/worker)
bee sex bias (queen/male)
C
bee sex bias (queen/male)
ant sex bias (queen/male)
D
fly sex bias (female/male)
shared queen
shared worker
abdominal caste bias

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). Sex bias in *D. melanogaster* is calculated based on whole bodies.

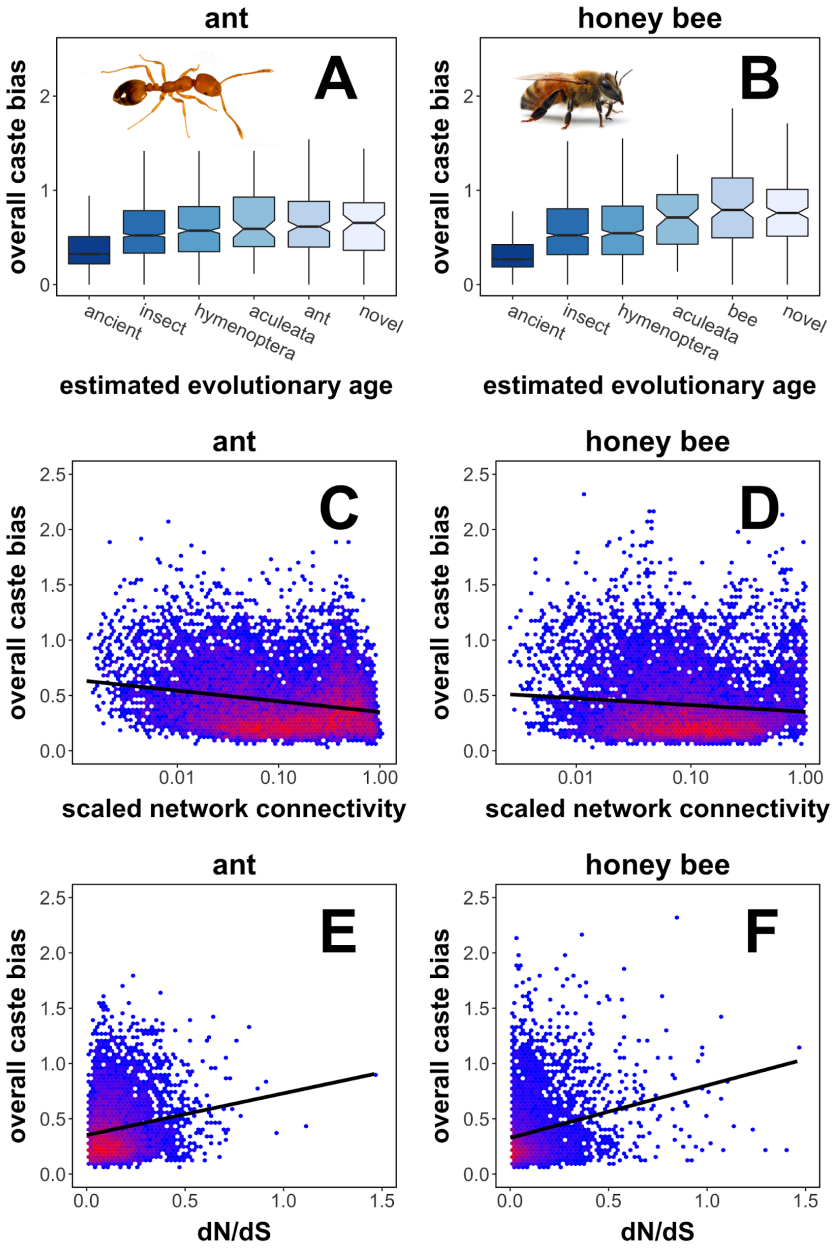

Fig. 4. Genes that exhibit more caste bias across tissues and developmental stages have younger estimated evolutionary ages (A,B) and tend to be loosely connected (C,D; ant: rho = -0.159, P < 0.001; honey bee: rho = -0.090, P < 0.001) and rapidly evolving (E,F; ant: rho = 0.157, P < 0.001; honey bee: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

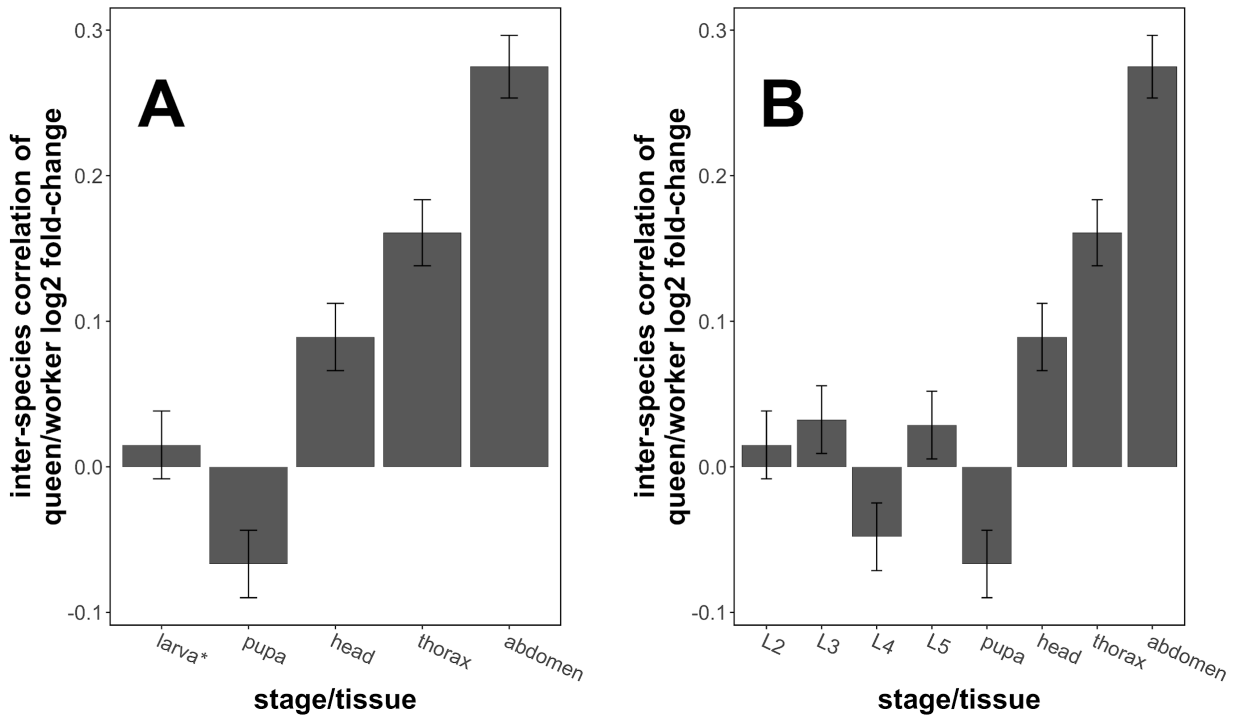

Fig. S1. Pearson correlation of log-fold change between queens and workers as measured at each stage or tissue in *M. pharaonis* and *A. mellifera* for each 1:1 ortholog (N = 7640). Error bars indicate Pearson correlation 95% confidence intervals. In (A), the category "larva*" represents differential expression across larval stages, while in (B) each larval stage (L2-L5) is plotted individually.

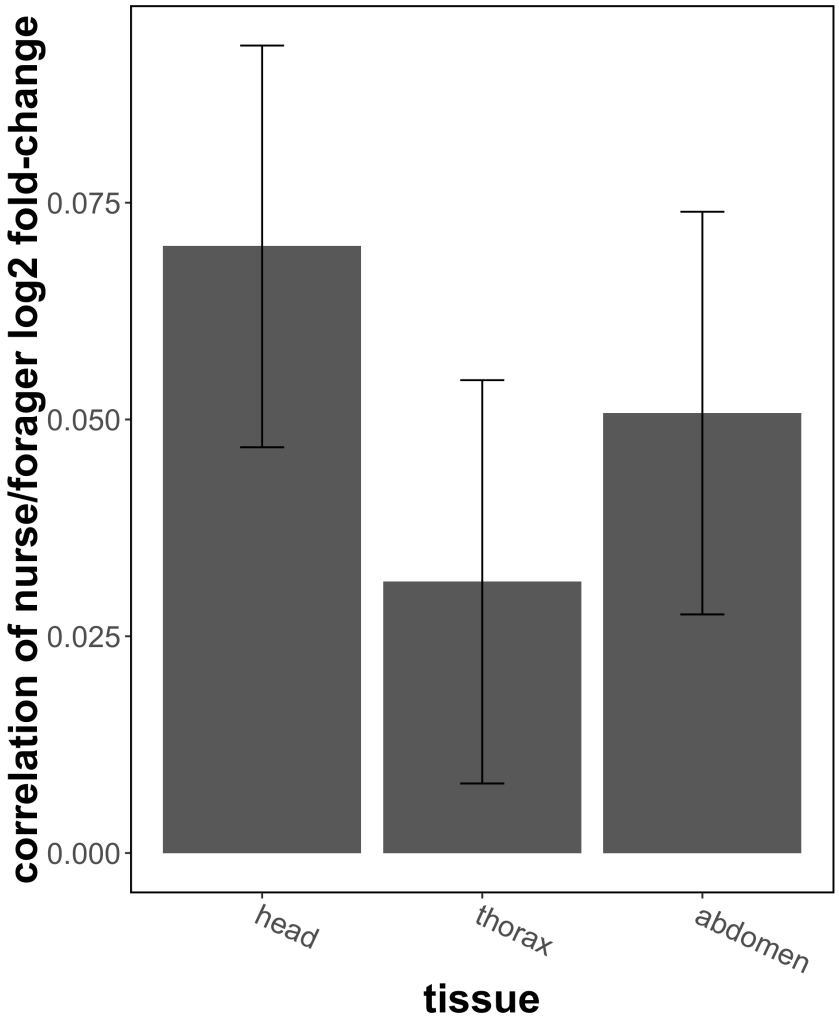

Fig. S2. Pearson correlation of log-fold change between nurses and foragers as measured in each tissue in *M. pharaonis* and *A. mellifera* for each 1:1 ortholog (N = 7640). Error bars indicate Pearson correlation 95% confidence intervals.

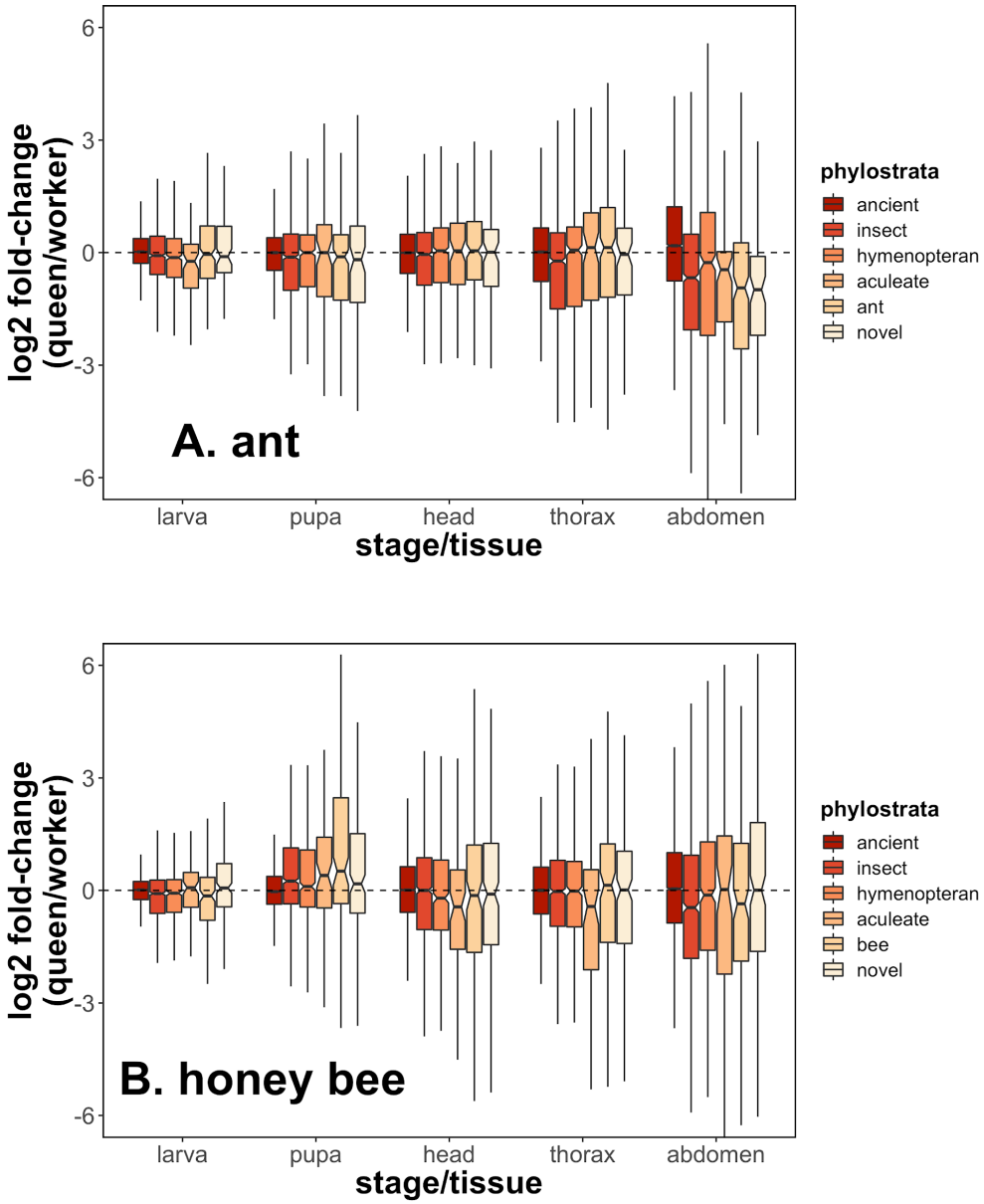

Fig. S3. Log_2 fold-change at each stage/tissue in each phylostrata category. Positive values indicating higher expression in queens compared to workers. Log_2 fold-change has been adjusted relative to the median value at that stage/tissue, in order to compare across tests. "Ancient" genes indicate any genes shared beyond insects (i.e. with vertebrates).

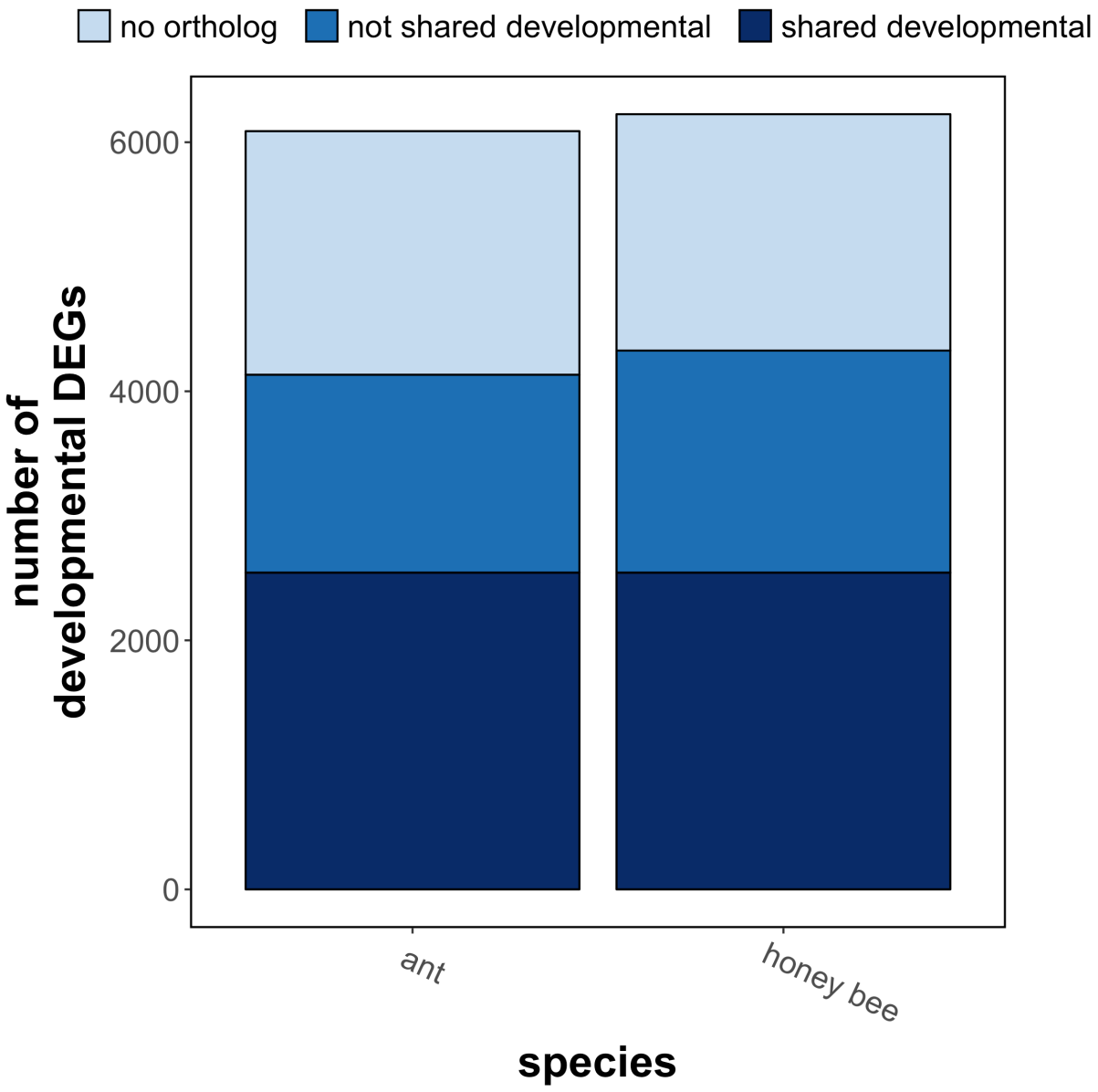

Fig. S4. Number of developmental differentially expressed genes in ants and in honey bees. "No ortholog" refers to genes for which no 1:1 ortholog exists (either due to apparent duplication or complete lack or orthology), "not shared developmental" refers to genes for which 1:1 orthologs can be identified but are only differentially expressed between developmental stages in one species, and "shared developmental" refers to genes for which 1:1 orthologs are differentially expressed between developmental stages in both species.

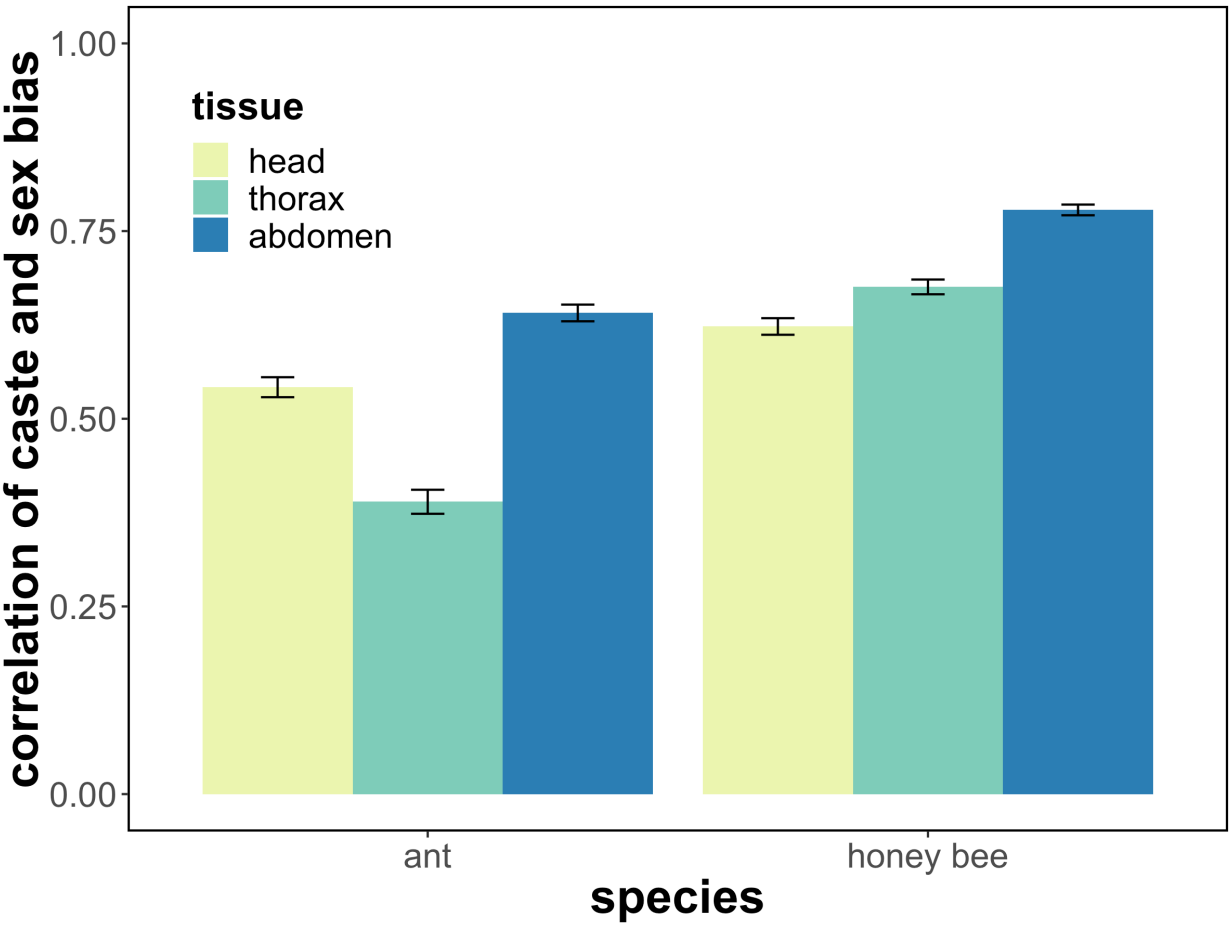

Fig. S5. Pearson correlation of caste (queen/worker) and sex (queen/male) expression bias in ants and honey bees. Error bars represent Pearson correlation 95% confidence intervals. Correlations are significant in all cases ($P < 0.001$), but abdominal correlations are strongest.

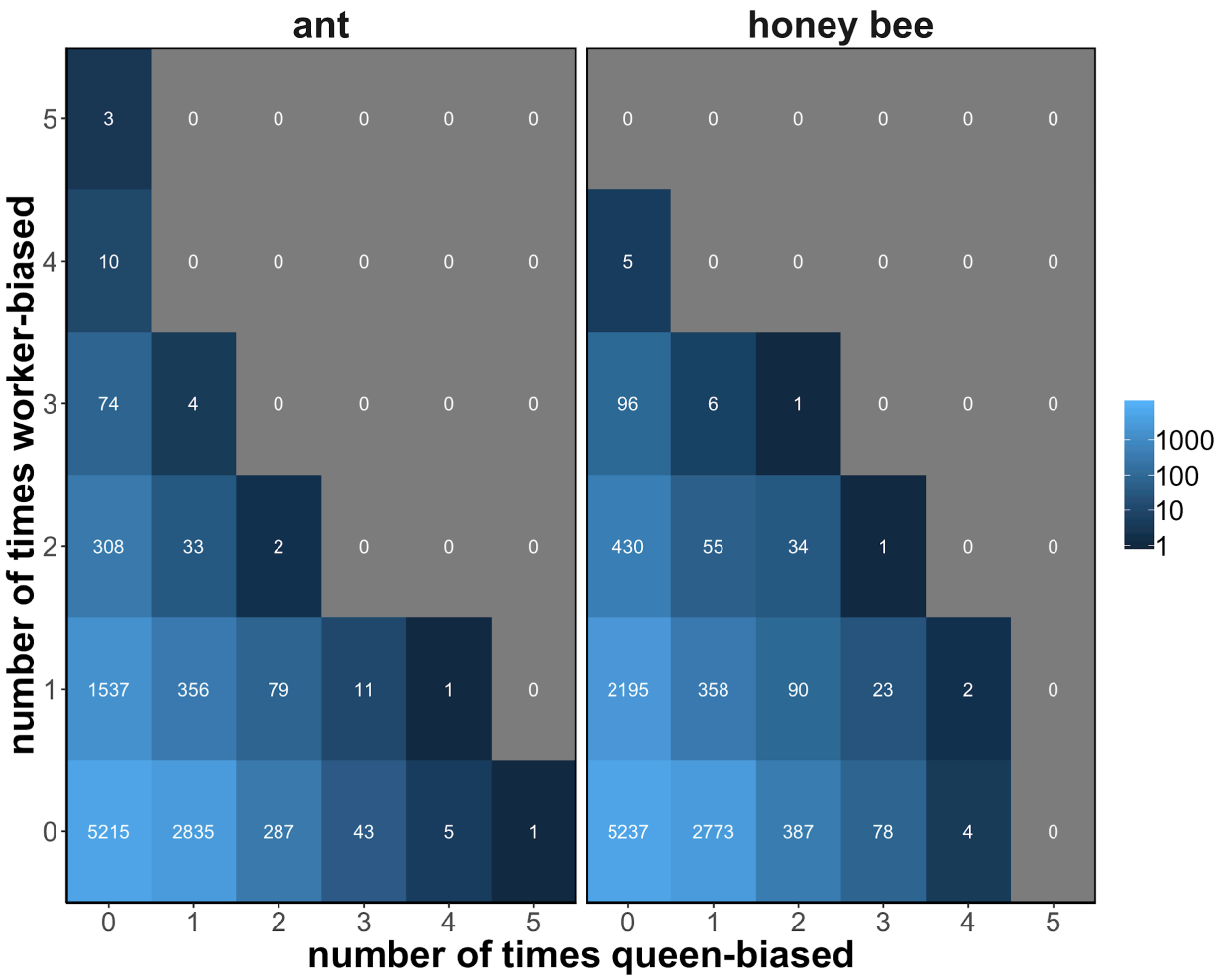

Fig. S6. Number of times each gene is upregulated in queen and workers across all comparisons (larva, pupa, and adult head, thorax, and abdomen). Color brightness is logarithmically proportional to the number of genes in each cell. N = 10804 genes for ants, N = 11775 genes for honey bees.

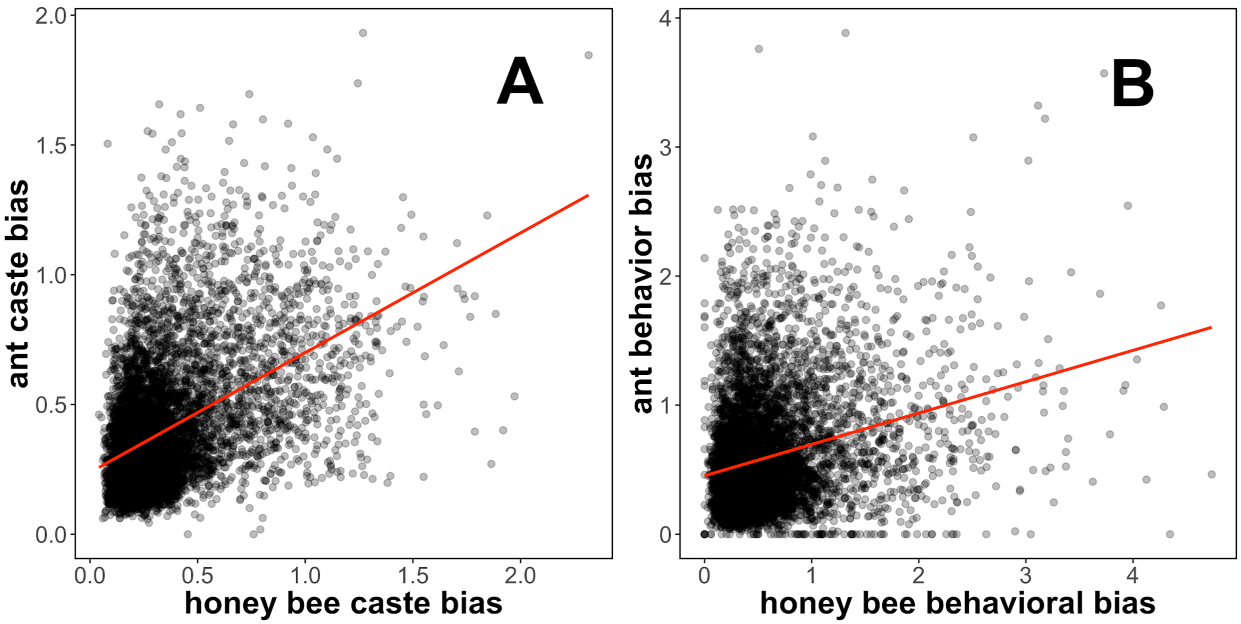

Fig. S7. Overall caste bias (A) and overall behavior bias (B) is correlated between ants and honey bees. "Overall" bias refers to the Euclidean distance of all $\log_2$ fold-change values (queens/worker for caste, nurses/foragers for behavior). The red line is the trendline of a linear model; Spearman correlation $P < 0.001$ in all cases.

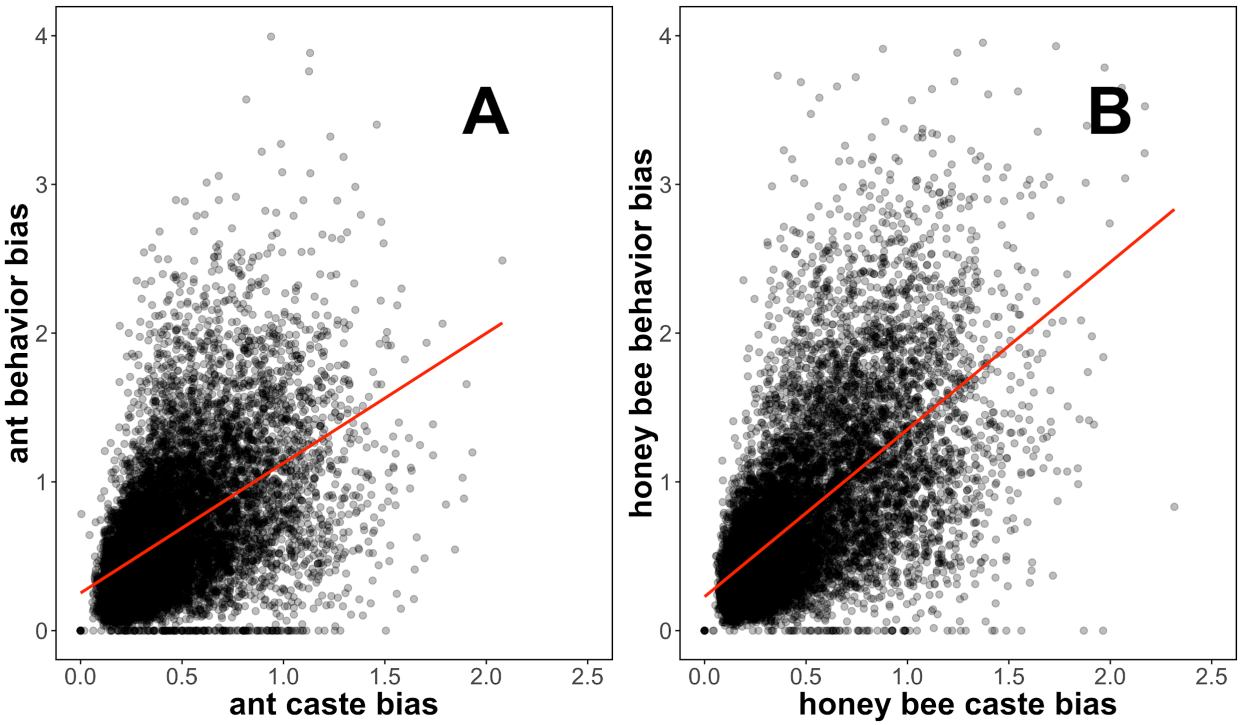

Fig. S8. Overall caste bias and overall behavior bias were correlated within A) ants and B) honey bees. "Overall" bias refers to the Euclidean distance of all $\log_2$ fold-change values (queens/worker for caste, nurses/foragers for behavior). The red line is the trendline of a linear model; Spearman correlation $P < 0.001$ in all cases.

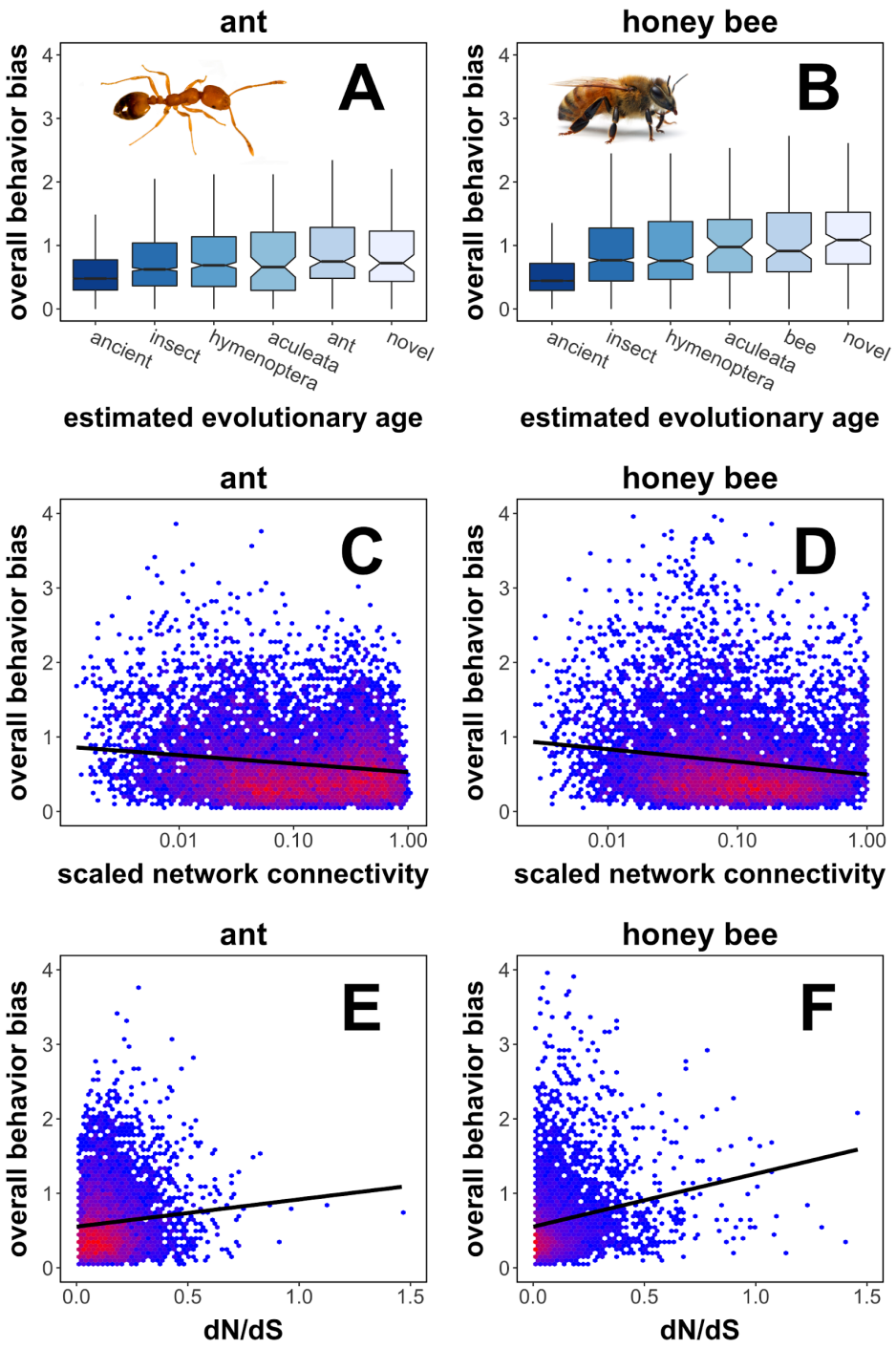

Fig. S9. Genes that exhibit more behavior bias across tissues have younger estimated evolutionary ages (A,B) and tend to be loosely connected (C,D; ant: rho = -0.099, P < 0.001; honey bee: rho = -0.157, P < 0.001) and rapidly evolving (E,F; ant: rho = 0.079, P < 0.001; honey bee: rho = 0.226, P < 0.001). "Overall behavior bias" combines nurse forager $\log_2$ fold-change values across all adult body segments. Connectivity is calculated using all samples and genes and scaled proportionally to the highest value.

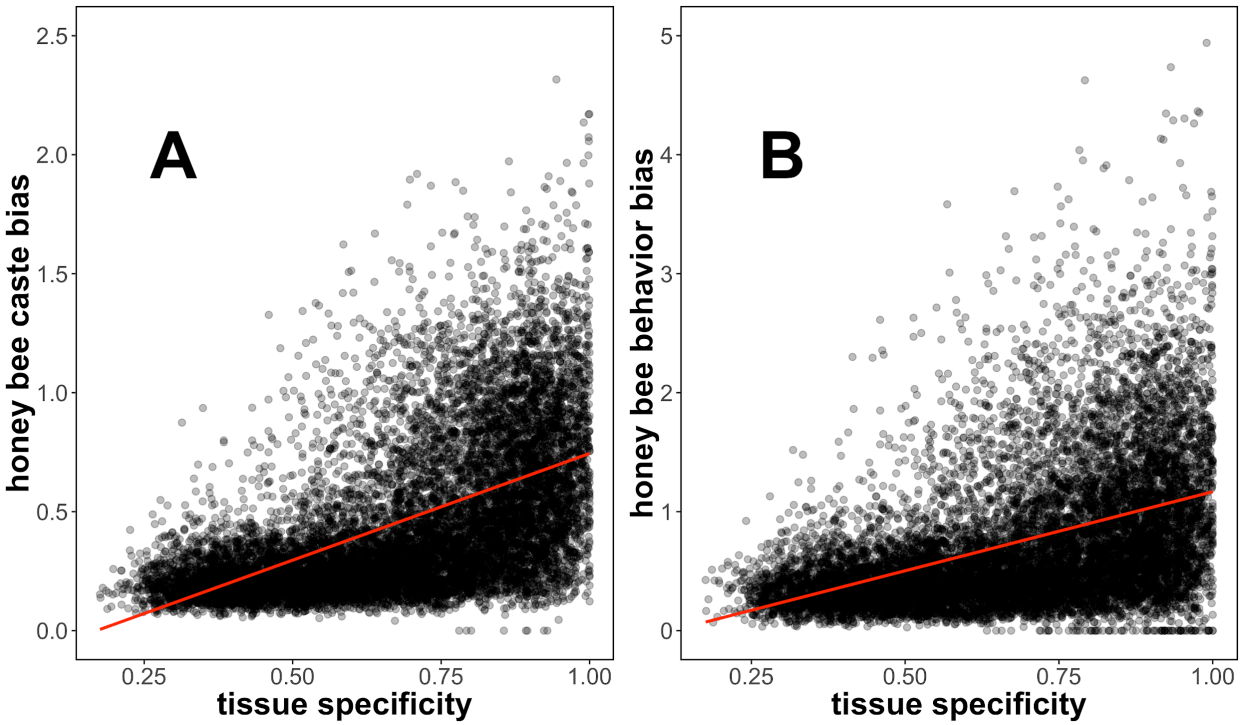

Fig. S10. Genes exhibiting more behavior bias tend to be tissue-specific. There was a positive correlation (Spearman correlation, $P < 0.001$ in each case) between caste/behavior bias and tissue specificity, where tissue specificity (□) is estimated using data from 12 honey bee tissues. □ = 1 indicates a genes is expressed in only one tissue, while lower values indicate genes are more ubiquitously (i.e. evenly) expressed across tissues. The red line is the trendline of a linear model.

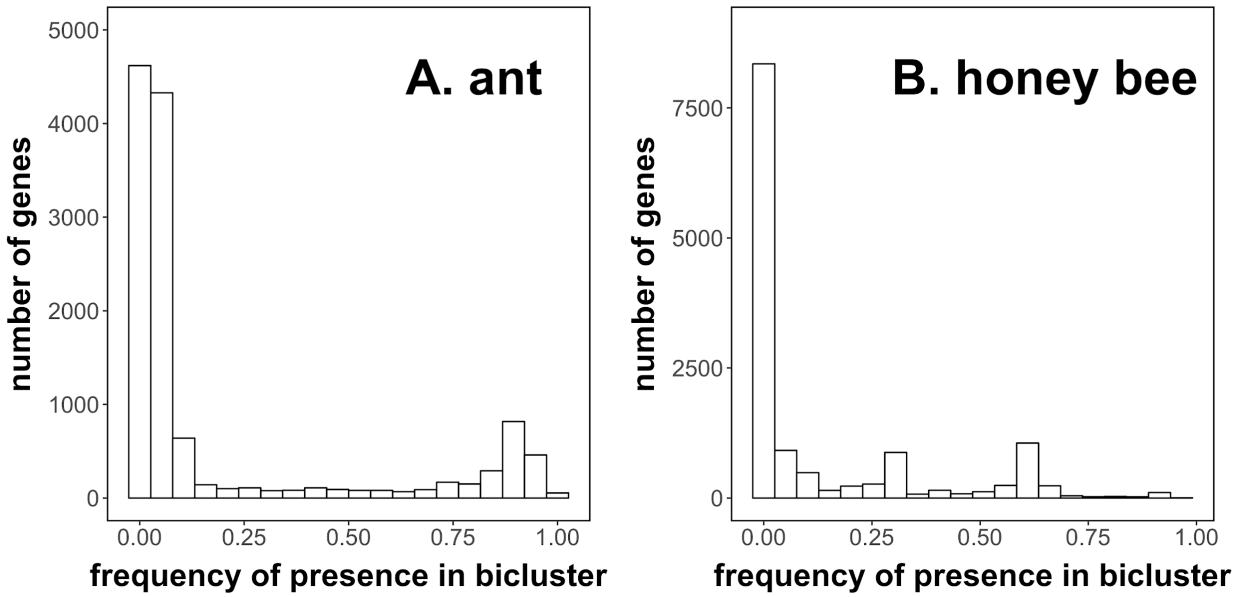

Fig. S11. Histogram of the frequency with which genes were placed in the queen abdomen bicluster (out of 1000 runs). Plaid biclustering is a non-deterministic process, so different sets of genes can be present in each run. In ants (A), 1039 genes were present in >90% of queen abdomen biclusters and retained 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.

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 queen	3
ant	adult	thorax	queen	virgin queen	3
ant	adult	abdomen	queen	virgin queen	3
ant	adult	head	queen	mated queen	3
ant	adult	thorax	queen	mated queen	3
ant	adult	abdomen	queen	mated queen	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 queen	3
honey bee	adult	thorax	queen	virgin queen	3
honey bee	adult	abdomen	queen	virgin queen	3
honey bee	adult	head	queen	mated queen	3
honey bee	adult	thorax	queen	mated queen	3
honey bee	adult	abdomen	queen	mated queen	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

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 (101). 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 (97). 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 (128). Sex and caste are both known in *A. mellifera* larvae, as individuals are reared in separate cells (16).

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:0007264	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	0.00297	pupa
GO:0046460	neutral lipid biosynthetic process	0.00297	pupa
GO:0046463	acylglycerol biosynthetic process	0.00297	pupa
GO:0072525	pyridine-containing compound biosynthetic process	0.0018	head
GO:0001704	formation of primary germ layer	0.0042	head
GO:0002098	tRNA 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.

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-rRNA	0.00126	abdomen
GO:0007436	larval salivary gland morphogenesis	0.00205	abdomen

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.

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 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.

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 developmental pigmentation	0.0056	abdomen
GO:0043474	pigment metabolic process involved in pigmentation	0.0056	abdomen

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	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-rRNA	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.

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
LOC105840891	5.777	0.973	Acidic repeat-containing protein
LOC105838786	5.389	0.824	Spondin-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-acetylglucosamine transferase and deubiquitinase 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
LOC105833898	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
LOC105831777	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

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
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	nucleoredoxin-like
LOC725841	6.339	0.897	hyaluronan mediated motility receptor
LOC100577382	6.272	0.896	targeting protein for Xklp2 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 maelstrom 2
LOC412031	5.811	0.930	S-phase kinase-associated protein 2
LOC102656846	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 gxcJ-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 Smaug homolog 1
LOC552725	3.868	0.848	N-acetylglucosamine-1-phosphotransferase subunits alpha/beta

Table S9. Hub genes of the queen abdominal module in honey bees. 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.

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).

Species	NCBI Taxonomy ID
Acromyrmex echinatior	103372
Atta cephalotes	12957
Atta colombica	520822
Camponotus floridanus	104421
Cardiocondyla obscurior	286306
Monomorium pharaonis	307658
Linepithema humile	83485
Lasius niger	67767
Harpegnathos saltator	610380
Dinoponera quadriceps	609295
Cyphomyrmex costatus	456900
Ooceraea biroi	2015173
Pogonomyrmex barbatus	144034
Pseudomyrmex gracilis	219809
Solenopsis invicta	13686
Trachymyrmex septentrionalis	34720
Trachymyrmex cornetzi	471704
Trachymyrmex zeteki	64791
Vollenhovia emeryi	411798
Wasmannia auropunctata	64793
Temnothorax curvispinosus	300111
Apis cerana	7461
Apis dorsata	7462
Apis florea	7463
Apis mellifera	7460
Bombus impatiens	132113
Bombus terrestris	30195
Melipona quadrifasciata	166423
Eufriesea mexicana	516756
Ceratina calcarata	156304
Megachile rotundata	143995
Habropoda laboriosa	597456
Dufourea novaeangliae	178035
Polistes canadensis	91411
Polistes dominula	743375
Ceratosolen solmsi	142686
Ceratosolen solmsi marchali	326594
Copidosoma floridanum	29053
Fopius arisanus	64838
Microplitis demolitor	69319
Nasonia vitripennis	7425
Trichogramma pretiosum	7493
Trichomalopsis sarcophagae	543379
Diachasma alloeum	454923
Orussus abietinus	222816
Athalia rosae	37344
Cephus cinctus	211228
Neodiprion lecontei	441921
Pediculus humanus	121225
Drosophila melanogaster	7227
Aedes aegypti	7159
Bombyx mori	7091
Papilio machaon	76193
Anopheles gambiae	7165
Onthophagus taurus	166361
Tribolium castaneum	7070
Acyrthosiphon pisum	7029
Zootermopsis nevadensis	136037
Caenorhabditis elegans	6239
Hydra vulgaris	6087
Strongylocentrotus purpuratus	7668
Lottia gigantea	225164
Helobdella robusta	6412
Mus musculus	10090
Homo sapiens	9606
Xenopus tropicalis	8364
Latimeria chalumnae	7897
Danio rerio	7955

Table S12. List of species used for phylostratigraphy analysis, with the NCBI Taxonomy ID.