1 **Genes associated with ant social behavior show distinct transcriptional and evolutionary**

2 **patterns**

3

4 Alexander S. Mikheyev[1,2] & Timothy A. Linksvayer[2]*

5

6 1. Ecology and Evolution Unit, Okinawa Institute of Science and Technology, Onna-son, Kunigami-gun,

7 Okinawa, Japan.

8 2. Research School of Biology, Australian National University, Canberra, ACT 0200, Australia.

9 Email: alexander.mikheyev@oist.jp

10

11 3. Department of Biology, University of Pennsylvania, Philadelphia PA 19104, USA. Email:

12 tlinks@sas.upenn.edu

13 *corresponding author

14

15

16   Abstract:

17   Studies of the genetic basis and evolution of complex social behavior emphasize either conserved or

18   novel genes. To begin to reconcile these perspectives, we studied how the evolutionary conservation of

19   genes associated with social behavior depends on regulatory context, and whether genes associated

20   with social behavior exist in distinct regulatory and evolutionary contexts. We identified modules of co-

21   expressed genes associated with age-based division of labor between nurses and foragers in the ant

22   *Monomorium pharaonis*, and we studied the relationship between molecular evolution, connectivity,

23   and expression. Highly connected and expressed genes were more evolutionarily conserved, as

24   expected. However, compared to the rest of the genome, forager-upregulated genes were much more

25   highly connected and conserved, while nurse-upregulated genes were less connected and more

26   evolutionarily labile. Our results indicate that the genetic architecture of social behavior includes both

27   highly connected and conserved components as well as loosely connected and evolutionarily labile

28   components.

29

30

31

32

33

34

35

36

37

38   The main conclusion of a decade of sociogenomic research with a range of solitary and social

39   animal species is that highly conserved genes underpinning core physiological processes can also

40   influence behavioral state (Amdam et al. 2004; Amdam et al. 2006; Toth and Robinson 2007; Toth et al.

41   2007; Toth et al. 2010; Woodard et al. 2011; Woodard et al. 2014). For example, the insulin signaling

42   pathway, which mediates an organism's response to its internal nutritional state, also influences its

43   behavior (Ament et al. 2008). The *genetic toolkit hypothesis* and related hypotheses propose that a

44   conserved set of genes and gene pathways involved in core physiological processes such as metabolism

45   and reproduction has been repeatedly used in the evolution of complex social behavior in diverse

46   lineages (West-Eberhard 1996; Amdam et al. 2004; Amdam et al. 2006; Toth and Robinson 2007; Toth et

47   al. 2007). This hypothesis stems from findings in Evolutionary Developmental Biology that morphological

48   innovation in disparate lineages often involves the convergent use of a conserved set of genes (e.g., Hox

49   genes) (Carroll et al. 2001; Toth and Robinson 2007; Wilkins 2013).

50   However, social behavior and other social traits are commonly viewed as having unique genetic

51   features and evolutionary dynamics, including especially rapid evolution (West-Eberhard 1983; Tanaka

52   1996; Moore et al. 1997; Wolf et al. 1999; Nonacs 2011; Bailey and Moore 2012; Van Dyken and Wade

53   2012). Could the molecular mechanisms underlying social interactions (e.g., social signal production and

54   response) and social behavior, together with the process of social evolution result in distinct genetic

55   architectures for social traits compared with other traits? Recent comparative transcriptomic and

56   genomic studies find low overlap in genes associated with social behavior in different highly social

57   animals and instead highlight the importance of novel genes and rapid evolution of social traits (Johnson

58   and Tsutsui 2011; Ferreira et al. 2013; Simola et al. 2013; Wissler et al. 2013; Feldmeyer et al. 2014;

59   Harpur et al. 2014; Jasper et al. 2014; Sumner 2014), in accordance with recent studies emphasizing the

60   ubiquity of taxonomically restricted genes (Domazet-Loso and Tautz 2003; Khalturin et al. 2009; Tautz

61   and Domazet-Loso 2011). Perhaps social evolution does not consistently use sets of highly conserved

62   genes to the same degree as morphological evolution? The *novel social genes hypothesis* proposes that

63   genes underlying social behavior are often novel socially-acting genes or are genes with novel social

64   functions not present in solitary ancestors (Johnson and Linksvayer 2010; Johnson and Tsutsui 2011;

65   Sumner 2014).

66   Research supporting the genetic toolkit hypothesis has stressed the significant signal of highly

67   conserved genes affecting core physiological processes in transcriptomic datasets for social behavior

68   (Robinson et al. 2008; Toth et al. 2010; Fischman et al. 2011; Woodard et al. 2011; Toth et al. 2014;

69   Woodard et al. 2014). In contrast, research supporting the novel social genes hypothesis has stressed

70 the overall low proportional overlap of genes underlying social behavior in divergent lineages as well as

71 the apparently general low degree of transcriptomic and genomic conservation in divergent lineages

72 (Johnson and Tsutsui 2011; Ferreira et al. 2013; Simola et al. 2013; Wissler et al. 2013; Feldmeyer et al.

73 2014; Harpur et al. 2014; Jasper et al. 2014; Sumner 2014).

74   We sought to build on these previous results by examining how transcriptional regulatory

75 context influences evolutionary conservation for genes associated with ant social behavior, and further

76 whether genes associated with ant social behavior exist in distinct regulatory and selective contexts

77 compared to the rest of the genome. Research in a range of model organisms demonstrates that the

78 degree of a gene's connectivity to the rest of the regulatory network and its level of expression are often

79 negatively correlated with its rate of molecular evolution (Krylov et al. 2003; Hahn and Kern 2005;

80 Jovelin and Phillips 2009; Ramsay et al. 2009). For example, highly connected "hub" genes are often

81 highly expressed and evolutionarily conserved. Previous research has compared rates of molecular

82 evolution for genes associated with reproductive division of labor in social insects (Hunt et al. 2010;

83 Hunt et al. 2013; Harpur et al. 2014), as well as other conditionally-expressed phenotypes in other

84 organisms (Brisson and Nuzhdin 2008; Leichty et al. 2012; Purandare et al. 2014), indicating that genes

85 associated with the expression of worker traits experience elevated rates of molecular evolution.

86 However, the relationships among molecular evolution, connectivity, and expression have been little

87 explored in social insects, and are generally little understood for genes associated with social behavior.

88 As a result it is unclear if observed differences in rates of molecular evolution are caused by differences

89 in regulatory architecture, expression, or perhaps result from distinct evolutionary mechanisms such as

90 kin selection, which may operate differentially on genes associated with social behavior relative to the

91 rest of the genome (Linksvayer and Wade 2009; Hall and Goodisman 2012). We further sought to

92 identify modules of co-expressed genes that may be composed of both conserved and novel genes and

93 may contribute to the expression and evolution of social complexity.

94   We studied the genetic basis and evolution of a fundamental aspect of social insect behavior,

95 age-based division of labor (age polyethism). Age polyethism involves the progression of workers from

96 in-nest tasks such as nursing to outside-nest tasks such as foraging. Because age polyethism is a trait

97 expressed by the functionally sterile worker caste, it is expected to be shaped primarily through kin

98 selection (Hamilton 1964). While age polyethism plays a central role in the functioning of many eusocial

99 systems (Hölldobler and Wilson 2009), the molecular underpinnings have only been well studied in the

100 honey bee *Apis mellifera* (Whitfield et al. 2006; Ament et al. 2008; Chandrasekaran et al. 2011), so that

101 the genetic and evolutionary basis of age polyethism is not generally understood outside of honey bees.

102   We identified transcriptional modules of co-regulated genes associated with worker age polyethism in

103   the pharaoh ant *Monomorium pharaonis*; we identified the degree that these genes overlap with genes

104   involved in age polyethism in two other social insects (Alaux et al. 2009; Manfredini et al. 2014); and we

105   studied the relationship between expression level, connectivity and rates of molecular evolution at

106   these genes compared to the rest of the genome.

107

108   **Results**

109

110   *Behavioral analysis.* We tracked cohorts of age-marked workers and recorded their behavior

111   and location inside and outside the nest. In order to identify differentially-expressed genes associated

112   with age-based division of labor, we collected age-marked workers and workers observed performing

113   specific behaviors. The observed location of workers from different age classes changed with both nest

114   location and behavior (glm with quasipoisson errors and log link, both p < 0.01) (Figure 1, Figure 1—

115   figure supplements 1 and 2). In concordance with the expected pattern of age polyethism, the average

116   age of workers observed in the different locations increased as distance from the brood area increased

117   (Figure 1—figure supplement 2). Of the 15 behaviors observed more than 15 total times (Supplemental

118   file 1), the likelihood of observing workers performing the behaviors "nurse", "groom," "rest",

119   "trophallaxis", "walk", and "forage" depended on worker age (Figure 1A; glm with binomial errors and

120   logit link, all nominal p < 0.0002, α = 0.003, controlling for multiple testing). Nursing and foraging were

121   at the two extremes: the average age of workers observed nursing was 6.94 days and the average age of

122   workers observed foraging (i.e. in the act of collecting food) was 13.04 days. There appeared to be a

123   marked transition from nursing to foraging between 9 and 12 days of age (Figure 1A), with 75% of

124   nursing observations made for workers less than 10 days old and 75% of foraging observations made for

125   workers over 10 days old (Figure 1—figure supplement 1).

126

127   *Genome and transcriptome assembly.* There was a trade-off in the assemblies between N50 and overall

128   assembly lengths, as a function of kmer size. We chose k=69 as a compromise between these two

129   metrics, resulting in a scaffolded assembly of 284mb, with a N50 of 19.0kb. Although there is no *M.*

130   *pharaonis* genome size estimate, the assembly is in the range of genome sizes typical of other

131   myrmicine ants (Tsutsui et al. 2008). CEGMA analysis (Parra et al. 2009) found complete sequences for

132   92% of the ultra-conserved eukaryotic genes, and partial sequences for 97%. Most reads (97.6%) could

133   be re-mapped to the genome assembly, resulting in a coverage estimate of 40x. Cufflinks assembly

134 identified 22,385 transcribed loci. 74.9 ± 18% (median 85.1%) of the reads for each sample could be re-

135 mapped to predicted transcripts extracted from the reference. After the reads were re-mapped to the

136 assembled transcripts using the RSEM pipeline, each library had 10,602,832 ± 2,925,898 expected

137 counts.

138

139 *Gene expression differences associated with worker behavior.* The complete analysis of gene expression

140 data, including R code and output, is available in the Supplemental file 2, and it is summarized below.

141 We wished to examine which of the four worker behavioral samples (nursing larvae, foraging, grooming

142 larvae, and worker-worker trophallaxis [i.e. exchanging liquid food]) had distinct expression profiles *vs.*

143 all of the others. We used linear contrasts to determine the number of differentially expressed genes

144 between the focal behavioral category and the other behaviors. Of these contrasts, only foragers and

145 nurses had significantly different gene expression patterns, when compared to the rest, *i.e.*, there was

146 no evidence that workers engaged in grooming and trophallaxis had distinct transcriptional states.

147 Consequently, we focused subsequent analysis on nurse and forager behavioral categories, except in the

148 construction of the co-expression networks, where all behavioral category and age class samples were

149 used (see below). There were 1217 forager-upregulated, 1247 nurse-upregulated transcripts, and

150 14,907 transcripts that were not differentially expressed.

151

152 *Gene expression associated with age polyethism.* Qualitatively, gene expression patterns mirrored the

153 behavioral transition from nursing to foraging that we observed around day 10 (Figure 1A and 1B). To

154 quantify these observations, we used a supervised learning approach (K-nearest neighbor classifier or

155 KNN) to check whether genes differentially expressed in nurses and foragers could be used to

156 differentiate the age class data into two clusters. After the KNN was trained on nurse and forager

157 profiles, it clearly separated workers into two distinct classes based on age, assigning those younger

158 than 12 days into the nurse class, and the rest into the forager class (Supplemental file 2 pages 14-15),

159 suggesting a fairly discrete transcriptomic transition between the two behaviors.

160

161 *Gene expression conservation analyses.* The proportion of genes with identified orthologs in the fire ant

162 *Solenopsis invicta* differed between behavioral categories (Manfredini et al. 2014), with forager-

163 upregulated genes having a higher proportion (0.54) relative to nurse-upregulated (0.43) and non-

164 differentially expressed (0.43) (multiple comparison Kruskal-Wallis, $p < 0.05$). Similarly, the proportion of

165 genes with identified honey bee *Apis mellifera* orthologs was higher for forager-upregulated genes

166   (0.50), relative to nurse-upregulated (0.38) and non-differentially expressed genes (0.38) (multiple

167   comparison Kruskal-Wallis, p < 0.05) (note we used a less conservative BLAST threshold for the honey

168   bee so that the proportions of honey bee and fire ant orthologs are not directly comparable, see

169   Methods). Furthermore, approximately half of non-differentially expressed (0.51) and nurse-

170   upregulated (0.50) genes did not have orthologs identified in either the fire ant or honey bee genomes,

171   but this proportion was lower for forager-upregulated genes (0.39); correspondingly, the proportion of

172   forager-upregulated genes with orthologs identified from both fire ants and honey bees was higher

173   (0.43) compared to nurse-upregulated and non-differentially expressed genes (0.32) (X^2 = 71.42, df = 6,

174    $p < 10^{-13}$).

175   Genes previously detected as upregulated in nurses and foragers of *S. invicta* were more likely

176   to have identified *M. pharaonis* orthologs up-regulated in these contexts as well (p = 0.0022 and p =

177   0.040, respectively). However, the actual percentage of genes differentially expressed in the same

178   context in these two ant data sets was small: 3.8 % (47/1247) of nurse genes and 3.2 % (39/1217) of

179   forager genes; or if only considering genes with orthologs identified in both species, 8.6% (47/549) nurse

180   genes and 5.9% (39/657) forager genes. While there was low overlap in the lists of differentially-

181   expressed genes, there could still be stronger overlap in genome-wide expression profiles when

182   comparing nurse and forager samples between *S. invicta* and *M. pharaonis.* Thus, we estimated the

183   correlation in the change of expression between nurse and forager samples (i.e. log fold change)

184   between the *S. invicta* and *M. pharaonis* datasets for all genes with identifiable homologs. There was a

185   significant correlation in the change of expression for nurse and forager samples, but one that explained

186   only 2% of the variance (Spearman's rho = 0.14, 6324 genes, $p < 10^{-16}$).

187   In contrast to the fire ant and pharaoh ant comparison, previously identified forager- and nurse-

188   upregulated honey bee *Apis mellifera* genes (Alaux et al. 2009) were not more likely to have *M.*

189   *pharaonis* orthologs expressed in the same context (p = 0.99, p = 0.98, respectively), consistent with a

190   previous comparison between *S. invicta* and *A. mellifera* (Manfredini et al. 2014). The actual overlap in

191   honey bee and pharaoh ant gene lists was higher (71 nurse-upregulated genes and 46 forager-

192   upregulated genes) due to the less conservative BLAST threshold we used for identifying honey bee

193   orthologs, but the honey bee lists were also larger (Alaux et al. 2009) and the overlap was not

194   significant.

195

196   *Gene Ontology analysis*. Nurse-upregulated genes were strongly enriched for a range of GO terms

197   associated with metabolism (nearly 50 metabolism-related terms with $p < 10^{-5}$; Supplemental file 3).

198   Forager-upregulated genes had a more diffuse signal, being relatively more weakly enriched for various

199   GO terms, e.g., associated with signal transduction and gland morphogenesis. Forager-upregulated

200   genes showed a more consistent signal for underrepresented terms, e.g., GO terms associated with

201   metabolic processes and chromatin modification (Supplemental file 3).

202

203   *Modules inferred by weighted gene co-expression network analysis (WGCNA).* The number of modules

204   produced by WGCNA can vary based on several thresholding parameters, which we left as defaults

205   (Supplemental file 2, pages 20-21). These settings resulted in 14 co-expression modules, ranging in size

206   from 83 to 4218 genes (Figure 1C; Figure 1—figure supplement 3). A module's overall expression can be

207   characterized by its eigengene. Correlations between eigengenes and traits in the original data suggests

208   the involvement of corresponding modules in these traits. Eigengenes in two of the modules -- 1 and 14,

209   which contained the most nurse and forager genes, respectively -- were strongly correlated with worker

210   age, although in opposite directions, suggesting their role in aging and age-based division of labor (r = -

211   0.95, r = 0.91 and with FDR-adjusted p-values 0.0038, 0.023, respectively) (Supplemental file 2, page 24).

212   Other modules showed complex patterns of age and behavior specific expression, with most of them

213   showing a peak in expression once or twice during the lifetime of a worker (Supplemental file 2, page

214   26). Interestingly, most module eigengenes switched signs during the period between 9 and 12 days,

215   corresponding to the behavioral transition from nursing to foraging. In other words, there appeared to

216   be a major reprogramming step, where modules initially showing low expression became up-regulated,

217   while modules initially showing high expression were down-regulated.

218

219   Forager-upregulated genes were concentrated in just a few modules, with only two modules containing

220   more than 100 forager-upregulated genes (Figure 1 – figure supplement 3). In contrast, nurse-

221   upregulated genes were more widely distributed, with five modules having more than 100 nurse-

222   upregulated genes (Figure 1 – figure supplement 3). These five modules were mainly enriched for GO

223   terms associated with metabolism and development (Figure 1 – figure supplement 3; Supplemental file

224   4). Module 5, which contained 116 nurse-upregulated genes, was also enriched for terms associated

225   with female gonad development, which is surprising given that *M. pharaonis* workers lack ovaries and

226   are completely sterile. The modules containing forager-upregulated genes were enriched for a broad

227   range of GO terms, for example associated with regulation of signaling, development and neurogenesis,

228   and gene expression (Figure 1 – figure supplement 3; Supplemental file 4). The proportion of module

229   genes with identified *S. invicta* orthologs ranged from 0.28 to 0.53 (Figure 1 – figure supplement 3),

230   suggesting that in addition to being involved in different functions, the modules are composed of

231   different proportions of conserved and taxonomically-restricted genes.

232

233   *Relationship between gene behavioral category, expression level, connectivity, and evolutionary rate.*

234   Forager-upregulated genes were much more connected than nurse or non-differentially expressed

235   genes, while nurse-upregulated genes were less connected than non-differentially expressed genes

236   (Figure 2A) (multiple comparison Kruskal-Wallis, $p < 0.05$). There was a small but significant difference in

237   evolutionary rate dN/dS (Figure 2C), with nurse-upregulated genes evolving more rapidly than non-

238   differentially expressed genes (multiple comparison Kruskal-Wallis, $p < 0.05$). Nurse and forager genes

239   were also more highly expressed (Figure 2B) than non-differentially expressed genes (Kruskal-Wallis, $p <$

240   0.05), although this last comparison is likely biased because differential expression is more easily

241   detected in highly expressed genes.

242   Co-expression network connectivity and expression level were overall negatively associated with

243   evolutionary rate, such that highly connected and highly expressed genes had decreased rates of

244   molecular evolution (Figure 2D and 2E; evolutionary rate and connectivity, $r = -0.15$, $p < 2 \times 10^{-16}$;

245   evolutionary rate and expression, measured in terms of transcriptional abundance, fragments per

246   million reads mapped, FPKM, $r = -0.12$, $p < 2 \times 10^{-16}$); and connectivity and expression were positively

247   correlated ($r = 0.30$, $p < 2 \times 10^{-16}$. In a full model considering how a gene's rate of molecular evolution

248   depended on its gene expression level, network connectedness, and behavioral category, the largest

249   effects were main effects of expression ($z = -7.42$, $p = 1.29 \times 10^{-13}$) and connectivity ($z = -3.69$, $p =$

250   0.00023).

251   We also studied the effects of gene category (i.e. upregulated in nurses or foragers, or not

252   differentially expressed), expression level, and connectivity on whether a given *M. pharaonis* gene had

253   an identifiable fire ant *S. invicta* and honey bee *A. mellifera* orthologs. Overall, genes with orthologs in

254   the fire ant or honey bee had greater connectivity and expression (Figure 3, Figure 3 – supplement 1). In

255   considering a model with both main and interaction effects of behavioral category, expression level, and

256   connectivity, connectivity had the strongest effect (glm with quasibinomial residuals: $t = 24.5$, $p < 10^{-16}$,

257   for the presence of *S. invicta* orthologs; $t = 32.2$, $p < 10^{-16}$, for the presence of *A. mellifera* orthologs),

258   with more highly connected genes being more likely to have an ortholog. There were also much smaller

259   interaction effects indicating that nurse-upregulated genes had fewer orthologs than expected given

260   their connectivities (i.e. connectivity had a weaker effect on nurse-upregulated genes than other genes,

261   Figure 3 and Figure 3 – supplement 1; $t = -3.17$, $p = 0.0015$ for *S. invicta* orthologs; $t = -2.76$, $p = 0.0057$

262 for *A. mellifera* orthologs), and forager-upregulated genes had fewer orthologs than expected given

263 their expression (t = -2.33, p = 0.02 for *S. invicta* orthologs; t = -2.58, p = 0.0098 for *A. mellifera*

264 orthologs; Figure 3 and Figure 3 – supplement 1).

265

266 **Discussion**

267

268   Pharaoh ant workers showed a clearly defined age-based transition from nursing to foraging, in

269 terms of both behavioral and transcriptional patterns, with nurses and foragers having strongly

270 differentiated sets of upregulated genes (Figure 1). We recovered the commonly observed genome-wide

271 relationship between a gene's rate of molecular evolution, its expression level, and its network

272 connectivity (Krylov et al. 2003; Hahn and Kern 2005; Jovelin and Phillips 2009; Ramsay et al. 2009).

273 Specifically, the rate of molecular evolution (dN/dS) as well as the likelihood a gene had identified fire

274 ant and honey bee orthologs was negatively correlated with its expression level and connectivity within

275 co-expression networks, while expression and connectivity were positively correlated (Figures 2, 3). In

276 addition to these genome-wide patterns, nurse- and forager-upregulated genes had distinct regulatory

277 and evolutionary patterns relative to each other and to the rest of the transcriptome (Figures 2, 3). Most

278 strikingly, forager-upregulated genes were much more highly connected and correspondingly more

279 conserved, while nurse-upregulated genes were less connected, and more rapidly evolving and less

280 conserved.

281   Previous studies of the evolutionary genetic basis of social behavior have focused on the overlap

282 of genes lists associated with social traits in different lineages. We found significant but seemingly low (<

283 4%) overlap in lists of differentially expressed genes and the correlation in genome-wide expression

284 profiles (r = 0.14) when comparing gene expression in nurse and forager samples between the pharaoh

285 ant and fire ant, *S. invicta*. Such low overlap seems surprising, given that these two ants are in closely

286 related ant genera, having diverged on the order of 50 Mya (Ward et al. 2014). However, the

287 comparison is not perfect, given substantial differences between the two studies in methodology used

288 to characterize the behaviors, and in the technology used to measure gene expression (i.e. microarray

289 vs. RNA sequencing) (Manfredini et al. 2014). We did not find significant overlap between lists of honey

290 bee and pharaoh ant genes associated with age polyethism, consistent with results reported by the

291 earlier fire ant study (Manfredini et al. 2014). While we expected decreased overlap given that honey

292 bees and ants diverged longer ago, ~170 Mya (Ronquist et al. 2012), and represent independent origins

293 of eusociality, the ant-honey bee comparison is also more problematic because the honey bee data are

294 based on brain gene expression profiles whereas the fire ant and pharaoh ant data are based on whole

295 body gene expression profiles.

296   Past studies have often interpreted significant but similarly low overlap in lists of genes

297 associated with social behavior from different lineages as supporting the genetic toolkit hypothesis

298 (Toth et al. 2010; Toth et al. 2014; Woodard et al. 2014). In contrast, other authors have recently

299 interpreted low overlap as being consistent with the novel social genes hypothesis, which emphasizes

300 the importance of taxonomically-restricted genes (Ferreira et al. 2013; Feldmeyer et al. 2014; Sumner

301 2014). The contrasting emphasis of authors on either conserved or novel genes begs the question: what

302 degree of conservation in gene lists is necessary for confirmation or rejection of these two hypotheses?

303 For example, the fact that nurse-upregulated genes in *M. pharaonis* are more rapidly evolving than the

304 rest of the genome and that 50% of nurse-upregulated genes do not have identifiable fire ant or honey

305 bee orthologs suggests that novel genes may have important nurse-specific functions. At the same time,

306 the significant overlap of fire ant and pharaoh ant gene lists and the strong enrichment of nurse-

307 upregulated genes for gene ontology terms associated with metabolism and development suggests that

308 conserved genes involved in core physiological processes also play important roles in nurse function and

309 the evolution of division of labor. Thus, our results are generally consistent with both hypotheses. We

310 suggest that neither of these two hypotheses has yet been formulated in a way that is readily tested, in

311 part because it is unclear what precise genes are expected to be included or excluded from a genetic

312 toolkit (Wilkins 2013). Furthermore, these hypotheses are not mutually exclusive, since both conserved

313 and novel genes likely play roles in the evolution of all new traits (Johnson and Linksvayer 2010;

314 Woodard et al. 2011).

315   We suggest that shifting the focus, from lists of genes to modules of co-expressed genes in the

316 context of genome-wide transcriptional and evolutionary patterns, can help to elucidate how social

317 evolution has produced social complexity. In this way, one question we can ask is whether we see any

318 simple molecular signature of social evolution, for example due to kin selection? As *Monomorium* ant

319 workers are obligately sterile, all worker traits are expected to be shaped exclusively by indirect

320 selection (i.e. kin selection) (Hamilton 1964). All-else-equal, such indirect selection is weaker than direct

321 selection, proportional to relatedness (Hamilton 1964), and *a priori* is expected to produce relaxed

322 selective constraint and elevated rates of molecular evolution for all genes associated with worker traits

323 (Linksvayer and Wade 2009). Past studies have found different rates of molecular evolution for worker-

324 biased and queen-biased genes, with most studies finding that worker-biased genes are more rapidly

325 evolving (Ferreira et al. 2013; Feldmeyer et al. 2014; Harpur et al. 2014; but see Hunt et al. 2010). Some

326   researchers have interpreted different patterns between lineages as being consistent with simple kin

327   selection predictions based on differences in within-colony relatedness (Hall and Goodisman 2012), but

328   most studies have emphasized the association between conditional expression and relaxed selection

329   (Hunt et al. 2011; Hunt et al. 2013), as well as genes associated with worker traits simply experiencing

330   stronger positive selection (Hunt et al. 2010; Ferreira et al. 2013; Feldmeyer et al. 2014; Harpur et al.

331   2014). We observed weakly elevated rates of molecular evolution at nurse-upregulated genes compared

332   to the rest of the genome, but much more notable was the distinct connectivity and corresponding

333   differences in gene conservation for forager-upregulated genes relative to nurse-upregulated and non-

334   differentially expressed genes. These results suggest that social evolution does not just have simple

335   genome-wide effects such as relaxed effective selection associated with kin selection, but instead

336   shapes complex social traits while acting within general systems-level constraints imposed by regulatory

337   architecture.

338   The common perception that social evolution often involves rapid evolutionary dynamics (West-

339   Eberhard 1983; Tanaka 1996; Moore et al. 1997; Wolf et al. 1999; Nonacs 2011; Bailey and Moore 2012;

340   Van Dyken and Wade 2012) may result from the fact that genes influencing many key social traits are

341   not only conditionally expressed, but are also located peripherally within regulatory networks, and so

342   are relatively unconstrained. For example, we expect that traits associated with social signal production

343   (e.g., pheromone and glandular secretions) are often located peripherally within regulatory networks

344   and as a result may be evolutionarily labile (Jasper et al. 2014), as is the case more generally with

345   secreted proteins (Julenius and Pedersen 2006; Liao et al. 2010; Nogueira et al. 2012). More core and

346   conserved components are also certain to be important to the expression of these traits, but their

347   contribution to trait evolution may be minimized by virtue of the fact that they are highly connected.

348   These arguments suggest how both conserved, toolkit genes, as well as rapidly evolving and

349   taxonomically restricted novel genes, likely play important roles in the evolution of social novelty, with

350   novel genes being added peripherally to regulatory networks. Our results are consistent with this

351   interpretation, because *M. pharaonis* age-based division of labor seems to have a complex genetic basis

352   with some components that are highly connected and conserved, and other components that are more

353   loosely connected and evolutionarily labile.

354   Our findings that nurse-upregulated genes are more rapidly evolving and less conserved among

355   social insect lineages relative to forager-upregulated genes suggest that nurse traits have been a major

356   focus of evolutionary innovation between social insect lineages. This result seems surprising given that

357   foragers of different lineages experience diverse environments outside the nest compared to the

358 relatively constant within-nest environment experienced by nurses, and could be expected to

359 experience more diverse selective pressures. One explanation is that the physiological mechanisms

360 associated with metabolically costly foraging activities and older adult life (*M. pharaonis* workers usually

361 only live several weeks [Peacock and Baxter 1950], so that foragers which start right before their second

362 week of age may already be senescing) may be relatively conserved and simple. Nursing behavior,

363 occurring during very early adult life, may involve more diverse physiological and developmental

364 processes, and nursing itself may also involve more diverse behaviors and physiological processes, such

365 as food processing and the synthesis of glandular secretions that are fed to larvae. Perhaps the relatively

366 more complex genetic architecture (less tightly connected, involving more modules and diverse

367 processes) has served as less of a constraint and facilitated more evolutionary change for nurse-related

368 genes. If so, we predict that nurse-specific functions and functions for early adult life may be generally

369 more evolutionarily labile as well as more physiologically and behaviorally labile within and across

370 lineages than forager-specific functions. Note that this prediction is opposite the typical expectation that

371 genes acting early in development have more pleiotropic effects and are thus especially constrained

372 (Roux and Robinson-Rechavi 2008; Piasecka et al. 2013), but obligate sterility may in part release

373 workers from these constraints on the evolution of genes acting early in worker development.

374   We identified two discrete sets of genes with distinct genetic architecture associated with age-

375 based division of labor. The majority of forager-upregulated genes were contained within a single gene

376 module (module 14; Figure 1 – figure supplement 3) that was significantly positively associated with age.

377 Another module with expression negatively associated with age contained the largest number of nurse

378 genes, but nurse genes were also broadly spread out across a number of other modules with complex

379 expression patterns across age and behavioral groups. Interestingly, the modules differed in the

380 proportion of constituent genes which had identifiable *S. invicta* and *A. mellifera* orthologs, indicating

381 that modules vary in the degree to which they are composed of conserved genes and gene networks

382 versus rapidly evolving genes with unknown function. That said, the modules were enriched for various

383 gene ontology terms, providing some insight into their putative functional importance (Supplemental

384 file 4).

385   By explicitly studying regulatory architecture and inferring modules of tightly connected genes in

386 other species as well as *M. pharaonis*, it will be possible to further identify what network components

387 contribute to the expression of social traits, how rapidly these components are evolving within

388 populations, and how they have contributed to phenotypic differences between divergent lineages.

389 Building on the genetic toolkit conceptual framework, it will be possible to ask to what degree diverse

13

390   lineages repeatedly use the same modules, and importantly approaches already exist for quantifying

391   module overlap in the absence of functional information (Oldham et al. 2006; Langfelder et al. 2011).

392   Similarly, after finding non-significant overlap in lists of genes associated with queen- and worker-caste

393   development in paper wasps and honey bees, Berens et al. (2014) recently invoked a "looser" version of

394   the genetic toolkit hypothesis by examining the overlap of inferred functional enrichment of gene lists

395   (i.e. via gene ontology analysis). Focus on co-expressed modules may actually improve the feasibility of

396   inferring the function of co-expressed genes based on observed expression patterns together with

397   standard functional information inferred from the subset of conserved annotated genes with

398   identifiable orthologs from model systems. It will also be possible to determine the relative contribution

399   of conserved versus taxonomically restricted genes to co-expression modules.

400

401   **Methods**

402

403   *Colony setup.* Two replicate *M. pharaonis* observation colonies were established, each with ten queens,

404   approximately 4000 workers, and 1000 brood, representing a random subsample of a larger source

405   colony. Each colony was established from a separate source colony, which came from a stock of

406   approximately 40 colonies that have been repeatedly mixed across generations so that they are

407   genetically similar. Observation nests were constructed of two pieces of 5x15cm glass separated by

408   1.5mm thick plastic sheeting. Colonies were given water in cotton-plugged test tubes, 50% honey

409   solution, beef liver, egg yolk, and mealworms *ad libitum*, replaced twice a week. Colonies were

410   maintained at 27 C and 65% relative humidity in climate controlled rooms at the University of

411   Pennsylvania.

412   Every three days, 600 newly eclosed callow workers, which were inferred to be approximately 0-

413   1 days old, were collected from 8-10 stock colonies. These callow workers were briefly anesthetized with

414   CO_2 and individually paint marked on the gaster with a unique age cohort color dot using a Sharpie extra

415   fine oil based paint pen, and then 300 were added to each of the observation colonies. Five uniquely

416   marked age cohorts were thus added to the colonies on days 1, 4, 7, 10, and 13 of the study. Nestmate

417   recognition is at most weak and transient in *M. pharaonis* (Schmidt et al. 2010)*,* and callows in particular

418   are readily accepted. We also set up a camera to automatically take images of the nest areas of each

419   colony once every 20 minutes for the entire period of the study, although we do not further discuss

420   these images.

421   Previous literature indicates that *M. pharaonis* workers are expected to live 9-10 weeks

422   (Peacock and Baxter 1950), but our preliminary trials with our setup indicated that workers tend to die

423   or lose their paint marks after several weeks. We ran the study for 1 month, expecting to capture the

424   major age-based transitions in worker behavior (e.g., the nursing to foraging transition observed in

425   other species), but it is possible that we missed late behavioral transitions that occurred towards the

426   end of workers' lives. In practice, such late transitions are difficult to detect as sample size necessarily

427   declines as increasing numbers of workers die.

428

429   *Behavioral analyses.* A behavioral scan of each colony was completed once each day for the duration of

430   the month-long study by recording the instantaneous behavior and location observed for every visible

431   paint-marked worker. Each behavioral scan was performed at 20x magnification with a Nikon SMZ800

432   stereomicroscope. We recorded 30 distinct behaviors, but only 15 were observed more than 15 total

433   times during the study period (Supplemental file 1). We defined an individual as foraging if it was

434   observed on a food or water source or actually carrying food (i.e. foraging included the behaviors "on

435   honey", "on liver", "on water", or "carrying food"; Supplemental file 1). Each experimental colony

436   contained four identifiable locations that were redefined prior to each behavioral scan: brood area,

437   brood periphery, remaining nest area, and foraging area. The brood area was defined as the central area

438   within the nest containing all brood and queens (Edwards 1991). The nest periphery was defined as the

439   region directly adjacent to the brood area, where workers were dense in aggregation but not in contact

440   with any of the brood. The nest area was defined as the sparsely-occupied remainder of the space

441   within the nest, not including the brood area and nest periphery. The foraging area included all areas

442   outside of the nest. Analyses of behavioral data were conducted in R (www.r-project.org).

443

444   *Worker sampling, genomic DNA sequencing, mRNA amplification and RNA library preparation.* Every

445   three days, whole bodies of 5 individuals from each available uniquely paint marked age cohort were

446   collected, flash frozen in liquid nitrogen, and stored at -80°C. This sampling scheme resulted in 7 groups

447   of individuals of known age (0, 3, 6, 9, 12, 15, and 18+ days old). 20 individuals of each of these age

448   category were pooled for whole body RNA extraction for each of the two replicate observation colonies.

449   In addition, for each of the two replicate observation colonies, we collected and pooled 20 non-paint

450   marked workers in the act of the following five behaviors: nursing larvae, grooming larvae, engaged in

451   trophallaxis with other workers, foraging for protein (collecting egg, mealworm, or liver), and foraging

452   for carbohydrates (collecting honey solution). RNA was extracted from pools of worker samples of

453   known age or observed behavior using Qiagen RNeasy kits with standard protocols. RNA sequencing

454   libraries were constructed at the University of Arizona Genetics Core (UAGC) with RNA TruSeq library

455   construction kits following standard protocols. In total there were 24 libraries: 2 colony replicates × (7

456   age groups + 5 behavioral groups). RNA sequencing was conducted at the University of Arizona Genetics

457   Core on an Illumina HiSeq2000 with 100 bp paired ends reads, with 6 samples multiplexed per lane,

458   randomly distributed across four lanes. Sequences were post-processed by cutadapt (Martin 2011) to

459   remove Illumina adapter sequences and ConDeTri (Smeds and Künstner 2011) to remove low-quality

460   bases.

461

462   *Reference genome sequencing and assembly.* DNA from a single haploid male (183ng) was used to

463   prepare a TruSeq library, which was sequenced in multiplex on an Illumina HiSeq 2000, yielding

464   70,894,179 million 100 bp read pairs. Raw genomic reads were quality and adaptor trimmed using

465   ConDeTri and cutadapt (Martin 2011; Smeds and Künstner 2011), producing 57,002,951 read pairs and

466   8,361,560 single reads (12.3 Gb total). The assembly was carried out using ABYSS, with a range of kmers

467   from 53 to 91 (Simpson et al. 2009). We then chose the assembly with the longest N50 as the reference

468   for transcriptome assembly. Genome assembly quality was evaluated using the CEGMA pipeline (Parra

469   et al. 2009), and by re-mapping the paired end trimmed reads using bowtie2 (Langmead and Salzberg

470   2012).

471

472   *Reference based transcriptome assembly, annotation and differential gene expression analysis.* The

473   transcriptome was mapped to the reference using Tophat 2, and assembled into transcripts using

474   Cufflinks 2.1 (Roberts et al. 2011; Kim et al. 2013). Gene expression data was obtained by re-mapping

475   the transcript reads to the extracted transcripts using RSEM and calculating the expected counts at the

476   gene level (Li and Dewey 2011). When multiple isoforms of a single locus were found, only the longest

477   transcript was used for gene annotation. Assembled transcripts were annotated using BLASTX from the

478   non-redundant NCBI database with expectation values of $E = 10^{-5}$. These results were used to assign

479   Gene Ontology (GO) profiles with *Blast2go* (Conesa et al. 2005).

480

481   *Differential gene expression analysis and transcriptional network analysis.* Transcript counts were

482   filtered by abundance, removing those with less than 1 fragment per kilobase mapped (FPKM) in more

483   than half of the libraries (Mortazavi et al. 2008). Differential gene expression analysis was carried out in

484   edgeR, using a GLM fit to the count data and identifying differentially expressed genes using planned

485   linear contrasts (Robinson et al. 2010). In order to infer co-expression modules and gain an insight into

16

486   network structure of gene interactions, we performed a weighted gene co-expression network analysis

487   (WGCNA) on the count data (Langfelder and Horvath 2008). WGCNA was conducted on the entire

488   transcript set, after filtering out the low-abundance transcripts. This analysis relies on patterns of gene-

489   co expression, but has been shown to reconstruct protein-protein interaction networks with reasonable

490   accuracy (Zhao et al. 2010; Allen et al. 2012). We used total connectivity as a measure of gene

491   interaction strength, because it is not as sensitive to module assignments, and most likely reflects the

492   overall selective pressures acting on the gene, beyond those imposed by its role in age polyethism. As

493   with most gene expression analysis, WGCNA provides better estimates for highly abundant genes, and in

494   particular for genes showing variation in their expression levels. Consequently, low-abundance and

495   invariant genes will show

496   GO term enrichment analysis was performed using the R package GOstats (Falcon and Gentleman

497   2007). We report GO terms as enriched when $p < 0.05$.

498

499   *Evolutionary rate and gene expression conservation analyses.* Fire ant (*Solenopsis invicta*) orthologs for

500   each gene were determined using reciprocal best BLASTP, using cutoffs of 10^{-10}. This parameterization

501   allowed for high specificity, though at the cost of sensitivity, since paralogs were ignored (Chen et al.

502   2007). These results were used to predict the *M. pharaonis* coding sequence using ORFPredictor (Min et

503   al. 2005). We then computed the pairwise dN/dS ratios for each gene using the branch model in PAML

504   (v. 4.6). Using the list of differentially expressed genes in foragers *vs.* nest workers in the fire ant

505   (Manfredini et al. 2014), Fisher's exact tests were used to examine whether genes differentially

506   expressed in these categories of workers were more likely conserved, than expected by chance. We

507   repeated the analysis above using honey bee (*Apis mellifera*) genes, except that the BLAST cutoff was

508   lowered to 10^{-5} to increase the chance of identifying orthologs in the more divergent honey bee.

509   To initially study whether evolutionary rate (dN/dS), connectivity (kTotal), and expression

510   (FPKM) differed between behavioral categories (nurse-upregulated, forager-upregulated, and non

511   differentially expressed), we used a Kruskal-Wallis test, adjusted for multiple comparisons (kruskalmc

512   function in the R package pgirmess). Finally, to study the main and interaction effects of connectivity,

513   expression, and behavioral category on evolutionary rate, we used a linear model log transformed rate

514   as the dependent variable, log transformed connectivity and expression as continuous predictors, and

515   behavioral category as a categorical predictor.

516

517 *Statistical analysis.* Statistical analysis was performed with R. Means are presented ± their standard

518 deviations. P-value cutoffs of 0.05 were used throughout the analysis. In the case of differential gene

519 expression data analyses were corrected for multiple comparisons using the Benjamini-Hochberg (FDR)

520 procedure (Benjamini and Hochberg 1995).

521

522 *Code and data availability.* Scripts for the bioinformatic analyses, and a README explaining the

523 workflow can be found at https://github.com/mikheyev/monomorium-polyethism. Most of the

524 workflow and output is shown in Supplemental file 2. All behavioral and gene expression data, including

525 a MySQL database for the gene expression data have been deposited to Dryad,

526 doi:10.5061/dryad.cv0q3. Raw reads and the genome assembly are available at the DNA Data Bank of

527 Japan, DDBJ BioProject PRJDB3164.

528

529 **Acknowledgements**

530

536

537 **References**

538

539 Alaux, C., S. Sinha, L. Hasadsri, G. J. Hunt, E. Guzman-Novoa, G. DeGrandi-Hoffman, J. L.
540   Uribe-Rubio, B. R. Southey, S. Rodriguez-Zas, and G. E. Robinson. 2009. Honey bee
541   aggression supports a link between gene regulation and behavioral evolution.
542   Proceedings of the National Academy of Sciences of the United States of America
543   106:15400-15405.
544 Allen, J. D., Y. Xie, M. Chen, L. Girard, and G. H. Xiao. 2012. Comparing Statistical Methods for
545   Constructing Large Scale Gene Networks. PLoS ONE 7.
546 Amdam, G. V., A. Csondes, M. K. Fondrk, and R. E. Page. 2006. Complex social behaviour
547   derived from maternal reproductive traits. Nature 439:76-78.
548 Amdam, G. V., K. Norberg, M. K. Fondrk, and R. E. Page. 2004. Reproductive ground plan
549   may mediate colony-level selection effects on individual foraging behavior in honey
550   bees. Proceedings of the National Academy of Sciences of the United States of
551   America 101:11350-11355.

552 Ament, S. A., M. Corona, H. S. Pollock, and G. E. Robinson. 2008. Insulin signaling is involved
553   in the regulation of worker division of labor in honey bee colonies. Proceedings of
554   the National Academy of Sciences of the United States of America 105:4226-4231.
555 Bailey, N. W., and A. J. Moore. 2012. Runaway Sexual Selection without Genetic
556   Correlations: Social Environments and Flexible Mate Choice Initiate and Enhance
557   the Fisher Process. Evolution 66:2674-2684.
558 Benjamini, Y., and Y. Hochberg. 1995. Controlling the False Discovery Rate - a Practical and
559   Powerful Approach to Multiple Testing. J Roy Stat Soc B Met 57:289-300.
560 Berens, A. J., J. H. Hunt, and A. L. Toth. 2014. Comparative transcriptomics of convergent
561   evolution: different genes but conserved pathways underlie caste phenotypes across
562   lineagse of eusocial insects. Molecular Biology and Evolution doi:
563   10.1093/molbev/msu330.
564 Brisson, J. A., and S. V. Nuzhdin. 2008. Rarity of males in pea aphids results in mutational
565   decay. Science 319:58-58.
566 Carroll, S. B., J. K. Grenier, and S. D. Weatherbee. 2001. From DNA to diversity: molecular
567   genetics and the evolution of animal design. Blackwell Science, Malden, MA.
568 Chandrasekaran, S., S. A. Ament, J. A. Eddy, S. Rodriguez-Zas, B. R. Schatz, N. D. Price, and G.
569   E. Robinson. 2011. Behavior-specific changes in transcriptional modules lead to
570   distinct and predictable neurogenomic states. Proceedings of the National Academy
571   of Sciences of the United States of America 108:18020-18025.
572 Chen, F., A. J. Mackey, J. K. Vermunt, and D. S. Roos. 2007. Assessing Performance of
573   Orthology Detection Strategies Applied to Eukaryotic Genomes. PLoS ONE 2.
574 Conesa, A., S. Götz, J. M. García-Gómez, J. Terol, M. Talón, and M. Robles. 2005. Blast2GO: a
575   universal tool for annotation, visualization and analysis in functional genomics
576   research. Bioinformatics 21:3674-3676.
577 Domazet-Loso, T., and D. Tautz. 2003. An evolutionary analysis of orphan genes in
578   Drosophila. Genome Research 13:2213-2219.
579 Edwards, J. P. 1991. Caste Regulation in the Pharaohs Ant Monomorium-Pharaonis -
580   Recognition and Cannibalism of Sexual Brood by Workers. Physiological
581   Entomology 16:263-271.
582 Falcon, S., and R. Gentleman. 2007. Using GOstats to test gene lists for GO term association.
583   Bioinformatics 23:257-258.
584 Feldmeyer, B., D. Elsner, and S. Foitzik. 2014. Gene expression patterns associated with
585   caste and reproductive status in ants: worker-specific genes are more derived than
586   queen-specific ones. Molecular Ecology 23:151-161.
587 Ferreira, P. G., S. Patalano, R. Chauhan, R. Ffrench-Constant, T. Gabaldon, R. Guigo, and S.
588   Sumner. 2013. Transcriptome analyses of primitively eusocial wasps reveal novel
589   insights into the evolution of sociality and the origin of alternative phenotypes.
590   Genome Biol. 14.
591 Fischman, B. J., S. H. Woodard, and G. E. Robinson. 2011. Molecular evolutionary analyses of
592   insect societies. Proceedings of the National Academy of Sciences of the United
593   States of America 108:10847-10854.
594 Hahn, M. W., and A. D. Kern. 2005. Comparative genomics of centrality and essentiality in
595   three eukaryotic protein-interaction networks. Molecular Biology and Evolution
596   22:803-806.

597 Hall, D. W., and M. A. D. Goodisman. 2012. The Effects of Kin Selection on Rates of Molecular
598   Evolution in Social Insects. Evolution 66:2080-2093.
599 Hamilton, W. D. 1964. The genetical evolution of social behaviour, I. Journal of Theoretical
600   Biology 7:1-16.
601 Harpur, B. A., C. F. Kent, D. Molodtsova, J. M. D. Lebon, A. S. Alqarni, A. A. Owayss, and A.
602   Zayed. 2014. Population genomics of the honey bee reveals strong signatures of
603   positive selection on worker traits. Proceedings Of The National Academy Of
604   Sciences Of The United States Of America 111:2614-2619.
605 Hölldobler, B., and E. O. Wilson. 2009. The Superorganism: The Beauty, Elegance, and
606   Strangeness of Insect Societies. Norton, New York, New York.
607 Hunt, B. G., L. Ometto, L. Keller, and M. A. D. Goodisman. 2013. Evolution at Two Levels in
608   Fire Ants: The Relationship between Patterns of Gene Expression and Protein
609   Sequence Evolution. Molecular Biology and Evolution 30:263-271.
610 Hunt, B. G., L. Ometto, Y. Wurm, D. Shoemaker, S. V. Yi, L. Keller, and M. A. D. Goodisman.
611   2011. Relaxed selection is a precursor to the evolution of phenotypic plasticity.
612   Proceedings of the National Academy of Sciences of the United States of America
613   108:15936-15941.
614 Hunt, B. G., S. Wyder, N. Elango, J. H. Werren, E. M. Zdobnov, S. V. Yi, and M. A. D. Goodisman.
615   2010. Sociality is linked to rates of protein evolution in a highly social insect.
616   Molecular Biology and Evolution 27:497-500.
617 Jasper, W. C., T. A. Linksvayer, J. Atallah, D. Friedman, J. C. Chiu, and B. R. Johnson. 2014.
618   Large scale coding sequence change underlies the evolution of post-developmental
619   novelty in honey bees. Molecular Biology and Evolution 10.1093/molbev/msu292.
620 Johnson, B. R., and T. A. Linksvayer. 2010. Deconstructing the superorganism: social
621   physiology, groundplans, and sociogenomics. Quarterly Review of Biology 85:57-79.
622 Johnson, B. R., and N. D. Tsutsui. 2011. Taxonomically restricted genes are associated with
623   the evolution of sociality in the honey bee. BMC Genomics 12.
624 Jovelin, R., and P. C. Phillips. 2009. Evolutionary rates and centrality in the yeast gene
625   regulatory network. Genome Biol. 10.
626 Julenius, K., and A. G. Pedersen. 2006. Protein evolution is faster outside the cell. Molecular
627   Biology and Evolution 23:2039-2048.
628 Khalturin, K., G. Hemmrich, S. Fraune, R. Augustin, and T. C. G. Bosch. 2009. More than just
629   orphans: are taxonomically-restricted genes important in evolution? Trends in
630   Genetics 25:404-413.
631 Kim, D., G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, and S. L. Salzberg. 2013. TopHat2:
632   accurate alignment of transcriptomes in the presence of insertions, deletions and
633   gene fusions. Genome Biol 14:R36.
634 Krylov, D. M., Y. I. Wolf, I. B. Rogozin, and E. V. Koonin. 2003. Gene loss, protein sequence
635   divergence, gene dispensability, expression level, and interactivity are correlated in
636   eukaryotic evolution. Genome Research 13:2229-2235.
637 Langfelder, P., and S. Horvath. 2008. WGCNA: an R package for weighted correlation
638   network analysis. BMC Bioinformatics 9:559.
639 Langfelder, P., R. Luo, M. C. Oldham, and S. Horvath. 2011. Is My Network Module Preserved
640   and Reproducible? Plos Computational Biology 7.
641 Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nature
642   Methods 9:357-U354.

643   Leichty, A. R., D. W. Pfennig, C. D. Jones, and K. S. Pfennig. 2012. Relaxed Genetic Constraint
644   is Ancestral to the Evolution of Phenotypic Plasticity. Integr Comp Biol 52:16-30.
645   Li, B., and C. N. Dewey. 2011. RSEM: accurate transcript quantification from RNA-Seq data
646   with or without a reference genome. BMC Bioinformatics 12:323.
647   Liao, B. Y., M. P. Weng, and J. Z. Zhang. 2010. Impact of Extracellularity on the Evolutionary
648   Rate of Mammalian Proteins. Genome Biol Evol 2:39-43.
649   Linksvayer, T. A., and M. J. Wade. 2009. Genes with social effects are expected to harbor
650   more sequence variation within and between species. Evolution 63:1685-1696.
651   Manfredini, F., C. Lucas, M. Nicolas, L. Keller, D. Shoemaker, and C. M. Grozinger. 2014.
652   Molecular and social regulation of worker division of labour in fire ants. Molecular
653   Ecology 23:660-672.
654   Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing
655   reads. EMBnet. journal 17:pp. 10-12.
656   Moore, A. J., E. D. Brodie, III, and J. B. Wolf. 1997. Interacting phenotypes and the
657   evolutionary process .1. Direct and indirect genetic effects of social interactions.
658   Evolution 51:1352-1362.
659   Mortazavi, A., B. A. Williams, K. McCue, L. Schaeffer, and B. Wold. 2008. Mapping and
660   quantifying mammalian transcriptomes by RNA-Seq. Nature Publishing Group
661   5:621-628.
662   Nogueira, T., M. Touchon, and E. P. C. Rocha. 2012. Rapid Evolution of the Sequences and
663   Gene Repertoires of Secreted Proteins in Bacteria. PLoS ONE 7.
664   Nonacs, P. 2011. Kinship, greenbeards, and runaway social selection in the evolution of
665   social insect cooperation. Proceedings of the National Academy of Sciences of the
666   United States of America 108:10808-10815.
667   Oldham, M. C., S. Horvath, and D. H. Geschwind. 2006. Conservation and evolution of gene
668   colexpression networks in human and chimpanzee brains. Proceedings of the
669   National Academy of Sciences of the United States of America 103:17973-17978.
670   Parra, G., K. Bradnam, Z. M. Ning, T. Keane, and I. Korf. 2009. Assessing the gene space in
671   draft genomes. Nucleic Acids Research 37:289-297.
672   Peacock, A. D., and A. T. Baxter. 1950. Studies in Pharaoh's ant, *Monomorium pharaonis* (L.),
673   3: Life history. Entomologist's Monthly Magazine 86:171-178.
674   Piasecka, B., P. Lichocki, S. Moretti, S. Bergmann, and M. Robinson-Rechavi. 2013. The
675   Hourglass and the Early Conservation Models-Co-Existing Patterns of
676   Developmental Constraints in Vertebrates. Plos Genet 9.
677   Purandare, S. R., R. D. Bickel, J. Jaquiery, C. Rispe, and J. A. Brisson. 2014. Accelerated
678   Evolution of Morph-Biased Genes in Pea Aphids. Molecular Biology and Evolution
679   31:2073-2083.
680   Ramsay, H., L. H. Rieseberg, and K. Ritland. 2009. The Correlation of Evolutionary Rate with
681   Pathway Position in Plant Terpenoid Biosynthesis. Molecular Biology and Evolution
682   26:1045-1053.
683   Roberts, A., H. Pimentel, C. Trapnell, and L. Pachter. 2011. Identification of novel transcripts
684   in annotated genomes using RNA-Seq. Bioinformatics 27:2325-2329.
685   Robinson, G. E., R. D. Fernald, and D. F. Clayton. 2008. Genes and Social Behavior. Science
686   322:896-900.

687 Robinson, M. D., D. J. McCarthy, and G. K. Smyth. 2010. edgeR: a Bioconductor package for
688   differential expression analysis of digital gene expression data. Bioinformatics
689   26:139-140.
690 Ronquist, F., S. Klopfstein, L. Vilhelmsen, S. Schulmeister, D. L. Murray, and A. P. Rasnitsyn.
691   2012. A Total-Evidence Approach to Dating with Fossils, Applied to the Early
692   Radiation of the Hymenoptera. Syst Biol 61:973-999.
693 Roux, J., and M. Robinson-Rechavi. 2008. Developmental Constraints on Vertebrate Genome
694   Evolution. Plos Genet 4.
695 Schmidt, A. M., P. d'Ettorre, and J. S. Pedersen. 2010. Low levels of nestmate discrimination
696   despite high genetic differentiation in the invasive pharaoh ant. Front Zool 7.
697 Simola, D. F., L. Wissler, G. Donahue, R. M. Waterhouse, M. Helmkampf, J. Roux, S. Nygaard,
698   K. M. Glastad, D. E. Hagen, L. Viljakainen, J. T. Reese, B. G. Hunt, D. Graur, E. Elhaik, E.
699   V. Kriventseva, J. Y. Wen, B. J. Parker, E. Cash, E. Privman, C. P. Childers, M. C. Munoz-
700   Torres, J. J. Boomsma, E. Bornberg-Bauer, C. R. Currie, C. G. Elsik, G. Suen, M. A. D.
701   Goodisman, L. Keller, J. Liebig, A. Rawls, D. Reinberg, C. D. Smith, C. R. Smith, N.
702   Tsutsui, Y. Wurm, E. M. Zdobnov, S. L. Berger, and J. Gadau. 2013. Social insect
703   genomes exhibit dramatic evolution in gene composition and regulation while
704   preserving regulatory features linked to sociality. Genome Research 23:1235-1247.
705 Simpson, J. T., K. Wong, S. D. Jackman, J. E. Schein, S. J. M. Jones, and İ. Birol. 2009. ABySS: a
706   parallel assembler for short read sequence data. Genome Research 19:1117-1123.
707 Smeds, L., and A. Künstner. 2011. ConDeTri - a content dependent read trimmer for
708   Illumina data. PLoS ONE 6:e26314.
709 Sumner, S. 2014. The importance of genomic novelty in social evolution. Molecular Ecology
710   23:26-28.
711 Tanaka, Y. 1996. Social selection and the evolution of animal signals. Evolution 50:512-523.
712 Tautz, D., and T. Domazet-Loso. 2011. The evolutionary origin of orphan genes. Nature
713   Reviews Genetics 12:692-702.
714 Toth, A. L., and G. E. Robinson. 2007. Evo-devo and the evolution of social behavior. Trends
715   in Genetics 23:334-341.
716 Toth, A. L., J. F. Tooker, S. Radhakrishnan, R. Minard, M. T. Henshaw, and C. M. Grozinger.
717   2014. Shared genes related to aggression, rather than chemical communication, are
718   associated with reproductive dominance in paper wasps (Polistes metricus). BMC
719   Genomics 15.
720 Toth, A. L., K. Varala, M. T. Henshaw, S. L. Rodriguez-Zas, M. E. Hudson, and G. E. Robinson.
721   2010. Brain transcriptomic analysis in paper wasps identifies genes associated with
722   behaviour across social insect lineages. P R Soc B 277:2139-2148.
723 Toth, A. L., K. Varala, T. C. Newman, F. E. Miguez, S. K. Hutchison, D. A. Willoughby, J. F.
724   Simons, M. Egholm, J. H. Hunt, M. E. Hudson, and G. E. Robinson. 2007. Wasp gene
725   expression supports an evolutionary link between maternal behavior and
726   eusociality. Science 318:441-444.
727 Tsutsui, N. D., A. V. Suarez, J. C. Spagna, and J. S. Johnston. 2008. The evolution of genome
728   size in ants. BMC Evolutionary Biology 8:64.
729 Van Dyken, J. D., and M. J. Wade. 2012. Origins of Altruism Diversity Ii: Runaway
730   Coevolution of Altruistic Strategies Via "Reciprocal Niche Construction". Evolution
731   66:2498-2513.

732 Ward, P. S., S. G. Brady, B. L. Fisher, and T. R. Schultz. 2014. The evolution of myrmicine
733   ants: phylogeny and biogeography of a hyperdiverse ant clade (Hymenoptera:
734   Formicidae). Syst Entomol DOI: 10.1111/syen.12090:1-21.
735 West-Eberhard, M. J. 1983. Sexual selection, social competition, and speciation. Quarterly
736   Review of Biology 58:155-183.
737 West-Eberhard, M. J. 1996. Wasp societies as microcosms for the study of development and
738   evolution. Pp. 290-317 *in* S. Turillazzi, and M. J. West-Eberhard, eds. Natural History
739   and Evolution of Paper Wasps. Oxford University Press, New York.
740 Whitfield, C. W., Y. Ben-Shahar, C. Brillet, I. Leoncini, D. Crauser, Y. LeConte, S. Rodriguez-
741   Zas, and G. E. Robinson. 2006. Genomic dissection of behavioral maturation in the
742   honey bee. Proceedings of the National Academy of Sciences of the United States of
743   America 103:16068-16075.
744 Wilkins, A. S. 2013. "The genetic tool-kit": the life-history of an important metaphor *in* J. T.
745   Streelman, ed. Advances in Evolutionary Developmental Biology. John Wiley & Sons,
746   Inc., Hoboken, NJ.
747 Wissler, L., J. Gadau, D. F. Simola, M. Helmkampf, and E. Bornberg-Bauer. 2013. Mechanisms
748   and Dynamics of Orphan Gene Emergence in Insect Genomes. Genome Biol Evol
749   5:439-455.
750 Wolf, J. B., E. D. Brodie, III, and A. J. Moore. 1999. Interacting phenotypes and the
751   evolutionary process. II. Selection resulting from social interactions. American
752   Naturalist 153:254-266.
753 Woodard, S. H., G. M. Bloch, M. R. Band, and G. E. Robinson. 2014. Molecular heterochrony
754   and the evolution of sociality in bumblebees (Bombus terrestris). P R Soc B 281.
755 Woodard, S. H., B. J. Fischman, A. Venkat, M. E. Hudson, K. Varala, S. A. Cameron, A. G. Clark,
756   and G. E. Robinson. 2011. Genes involved in convergent evolution of eusociality in
757   bees. Proceedings of the National Academy of Sciences of the United States of
758   America 108:7472-7477.
759 Zhao, W., P. Langfelder, T. Fuller, J. Dong, A. Li, and S. Hovarth. 2010. Weighted Gene
760   Coexpression Network Analysis: State of the Art. J Biopharm Stat 20:281-300.

761

762

763 **Figure legends.**

764 Figure 1. Behavioral and transcriptional changes associated with worker age and behavior.

765 Numbers along the y-axis represent ages of marked worker cohorts, starting at worker eclosion as an

766 adult (day 0). In all plots dark green represents greater values, while white represents lower values of

767 the measure being plotted. (A) Behavioral results. Workers showed an age-dependent progression of

768 activity, progressing from tasks such as nursing and grooming in the nest to outside tasks such as

769 walking and foraging.

770 (B) Heat map of expression levels over the course of worker aging (higher expression in darker green),

771 for 25 genes most differentially expressed between nurses and foragers. The red line separates the

772   samples classified as 'nurses' by K-nearest neighbor classification on the left, from 'foragers' on the

773   right, suggesting a distinct transition between the two categories.

774   (C) Correlation between patterns of expression in the 14 identified modules across worker age and

775   behavior. The colors of the boxes are scaled with the value of correlation coefficients, ranging from

776   white to dark green. On the right side of the heat map are the numbers of genes in each module and a

777   dendrogram showing the inferred relationships among modules. The modules show complex patterns of

778   expression, for example with some most highly expressed at age 0, some showing decreasing expression

779   over time, and some increasing expression over time.

780

781   Figure 1 – figure supplement 1. The behaviors performed by age-marked workers changed as the

782   workers aged, from nursing to foraging. Boxplots show the distribution of age in days for each behavior;

783   white diamonds and the printed number show the mean age for each behavior; and the number of

784   observations for each behavior are shown at the bottom of the graph.

785

786   Figure 1 – figure supplement 2. The location of age-marked workers also changed as the workers aged,

787   from the nest area over the brood to outside the nest. Boxplots show the distribution of age with the

788   mean and sample size for each category.

789

790   Figure 1 – figure supplement 3. The identified modules vary in expression pattern, composition of nurse-

791   upregulated and forager-upregulated genes, and the proportion of conserved genes with identified fire

792   ant orthologs. The total number of genes, number of nurse-upregulated genes, and forager-upregulated

793   genes are shown, along with the proportion of identified fire ant orthologs and prominent gene

794   ontology terms enriched for each module (see Supplemental file 4 for the full GO enrichment profiles for

795   each module). Both the module number and associated module color are shown on the left.

796

797   Figure 2. Connectivity, expression, and evolutionary rate for nurse-upregulated (blue), forager-

798   upregulated (red), and non-differentially expressed genes (grey). Overall, connectivity and expression

799   are positively correlated (F) and negatively associated with evolutionary rate (D and E), as expected. At

800   the same time, forager-upregulated genes are much more strongly connected while nurse-upregulated

801   genes are more loosely connected compared to non-differentially-expressed genes (A); Nurse-

802   upregulated genes have a small but significant increase in evolutionary rate (C); and both forager- and

803   nurse-upregulated genes are more highly expressed than non-differentially expressed genes. The top

804   panels show results for all data, while the bottom panels show results only for genes with *S. invicta*

805   orthologs that had estimated evolutionary rates.

806

807   Figure 3. Genes with identified fire ant orthologs were more highly connected and expressed, but this

808   relationship also depended on whether the gene was nurse-upregulated (blue), forager-upregulated

809   (red), or non-differentially expressed (NDE, grey). As shown in Figure 2, forager-regulated genes were

810   much more highly connected, and overall, forager-upregulated genes had a higher proportion of

811   identified fire ant orthologs (0.54) relative to nurse-upregulated and non-differentially expressed genes

812   (0.43).

813

814   Figure 3 – figure supplement 1. Very similarly to Figure 3, genes with identified honey orthologs were

815   more highly connected and expressed, but this relationship also depended on whether the gene was

816   nurse-upregulated (blue), forager-upregulated (red), or non-differentially expressed (NDE, grey).

817   Forager-upregulated genes had a higher proportion of identified honey bee orthologs (0.50) relative to

818   nurse-upregulated and non-differentially expressed genes (0.38).

819

820   **Supplemental file captions**

821   Supplemental file 1. Table with a list of behaviors observed during the study; behavioral

822   code, name of the behavior, description of the behavior, the label for the behavior used in

823   the main text are shown, and the total number of observations for each behavior are

824   shown.

825

826   Supplemental file 2. A PDF of the full R script and abbreviated output with the analyses

827   included in the study.

828

829   Supplemental file 3. A table showing the results of Gene Ontology enrichment analysis for

830   the set of genes upregulated in foragers and nurses. The columns show: the GO Biological

831   Process ID, the P value, Odds Ratio, Expected Count, Observed Count, and total number of

832   genes in the data set matching the GO term, a description of the GO term, whether the term
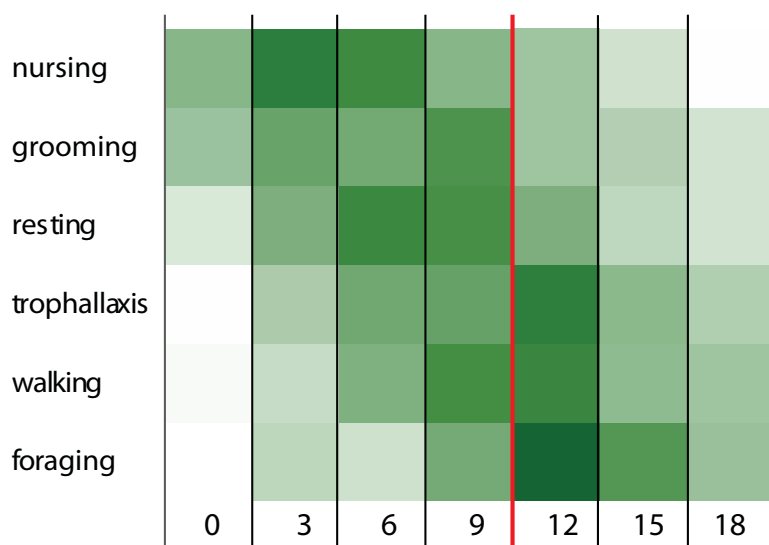
833   was enriched for forager-upregulated genes or nurse upregulated genes, and finally

834   whether the term is over- or underrepresented.

835   Supplemental file 4. A table showing the results of Gene Ontology enrichment analysis for

836   genes within each of the 14 identified co-expression modules. The columns show: the GO

837   Biological Process ID, the P value, Odds Ratio, Expected Count, Observed Count, and total

838   number of genes in the data set matching the GO term, a description of the GO term, and the

839   module number and module color in which the GO term was overrepresented.

840   Supplemental File 5 – Source R Markdown File for Supplementary File 2
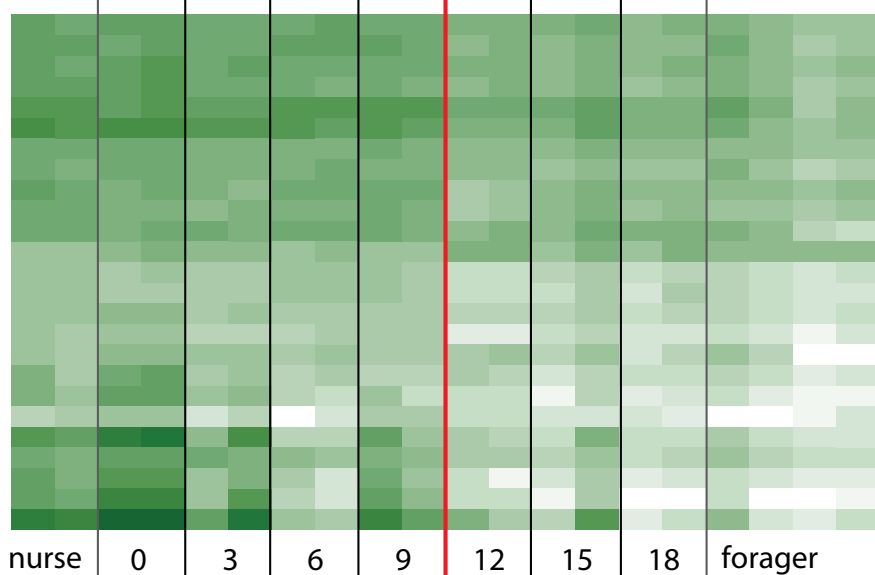
A

behavior

nursing							
grooming							
resting							
trophallaxis							
walking							
foraging							

0 3 6 9 12 15 18

B

gene expression

nurse 0 3 6 9 12 15 18 forager

C

gene modules

1		1059
2		398
3		1804
4		136
5		859
6		83
7		200
8		2405
9		694
10		4218
11		120
12		1373
13		403
14		3619

nurse 0 3 6 9 12 15 18 forager