



Evolutionary constraints shape caste-specific gene expression across 15 ant species

Author	Claire Morandin, Alexander S. Mikheyev, Jes Soe Pedersen, Heikki Helanterä
journal or publication title	Evolution
volume	71
number	5
page range	1273-1284
year	2017-03-06
Publisher	John Wiley and Sons
Rights	(C) 2017 The Author(s) This is the peer reviewed version of the following article: Morandin, C. , Mikheyev, A. S., Pedersen, J. S. and Helanterä, H. (2017), Evolutionary constraints shape caste specific gene expression across 15 ant species. Evolution, 71: 1273-1284. doi:10.1111/evo.13220, which has been published in final form at https://doi.org/10.1111/evo.13220 . This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.
Author's flag	author
URL	http://id.nii.ac.jp/1394/00000367/

doi: info:doi/10.1111/evo.13220

26 **Author contributions:**

27 Original idea CM, ASM, JSP, HH; Study design HH; Data collection CM, ASM, HH;
28 Data analyses CM; Manuscript preparation CM, HH. All authors read and approved
29 the final manuscript.

30

31 **Funding:**

32 This work was supported by the Academy of Finland (grant numbers 140990,135970,
33 273029 to H.H., and 252411 to the Centre of Excellence in Biological Interactions),
34 by Helsinki University, by the Kone Foundation (HH), by the Okinawa Institute of
35 Science and Technology Graduate University, and JSPS KAKENHI # 24770034 and
36 25221206 (to A.S.M.), and by the Danish National Research Foundation (grant
37 number DNRF57 to the Centre for Social Evolution).

38

39 **Acknowledgements:**

40 We are grateful to Steven D. Aird and Luke Holman for comments on the manuscript.

41

42

43 **Data availability**

44 The raw reads of the transcriptomes are publicly available in the DNA Data Bank of
45 Japan under bioproject ID PRJDB4088, sample accession numbers ID
46 SAMD00035735-SAMD00035834 (Morandin et al. 2016). All transcriptome
47 assemblies can be found on Fourmidable (Wurm et al. 2009). All data used in this
48 study can be found in Table S1.

49

50

51 **Abstract**

52 Development of polymorphic phenotypes from similar genomes requires gene
53 expression differences. However, little is known about how morph-specific gene
54 expression patterns vary on a broad phylogenetic scale. We hypothesize that evolution
55 of morph specific gene expression, and consequently morph specific phenotypic
56 evolution, may be constrained by gene essentiality and the amount of pleiotropic
57 constraints. Here, we use comparative transcriptomics of queen and worker morphs,
58 i.e. castes, from 15 ant species to understand the constraints of morph-biased gene
59 expression. In particular, we investigate how measures of evolutionary constraints at
60 the sequence level (expression level, connectivity, and number of gene ontology
61 terms) correlate with morph-biased expression. Our results show that genes indeed
62 vary in their potential to become morph-biased. The existence of genes that are
63 constrained in becoming caste-biased potentially limits the evolutionary decoupling of
64 the caste phenotypes, i.e. it might result in ‘caste load’ occasioning from antagonistic
65 fitness variation, similarly to sexually antagonistic fitness variation between males
66 and females. On the other hand, we suggest that genes under low constraints are
67 released from antagonistic variation and thus more likely to be co-opted for morph
68 specific use. Overall, our results suggest that the factors that affect sequence
69 evolutionary rates and evolution of plastic expression may largely overlap.

70

71

72

73

74

75

76 **Introduction**

77 Gene expression differences are the key to polyphenic development, such as
78 environmentally determined sex specific phenotypes (e.g. fish (Manousaki et al.
79 2014), reptiles (Yatsu et al. 2016)), feeding morphs (e.g. toad tadpoles (Leichty et al.
80 2012)) dispersal morphs (e.g. aphids (Braendle et al. 2006)), or social insect female
81 castes (Wilson 1971)). Given that polyphenic phenotypes are built from similar sets of
82 genes, with the exception of sex chromosomes, differential gene expression is
83 necessary for selection to work in a morph specific manner (Evans and Wheeler 1999;
84 Pereboom et al. 2005; Sumner et al. 2006; Parsch and Ellegren 2013; Perry et al.
85 2014). Thus, understanding the features of genes that are expressed in a morph-
86 specific manner (hereafter “morph-biased genes”) is crucial for understanding
87 evolution of polymorphism.

88

89 The increased availability of transcriptomic data has revealed the dual nature of
90 morph-biased genes (Mikheyev and Linksvayer 2015). On the one hand, morph-
91 biased genes are disproportionately likely to be genes that evolve fast, either under
92 relaxed purifying selection (Hunt et al. 2010, 2011; Leichty et al. 2012; Helanterä and
93 Uller 2014; Schrader et al. 2016), or positive selection (Devlin and Nagahama 2002;
94 Guidugli et al. 2005; Ellegren and Parsch 2007; Mank and Ellegren 2009; Harpur et
95 al. 2014; Yang et al. 2016). Fast evolution of morph-biased genes is also suggested by
96 overrepresentation of taxonomically restricted genes in morph-biased genes,
97 especially in social insects (Ferreira et al. 2013; Feldmeyer et al. 2014; Jasper et al.
98 2014; Sumner 2014). On the other hand, in line with the genetic toolkit hypothesis,
99 genes with conserved functions are thought to represent the shared genomic basis of
100 polyphenism; comparative studies suggest that such functions include core

101 physiological processes such as metabolism and reproduction (Swain and Lovell-
102 badge 1999; Nef et al. 2005; Toth and Robinson 2007; Toth et al. 2010; Fischman et
103 al. 2011; Woodard et al. 2011; Berens et al. 2014). In general, genes with conserved
104 or essential function (Richards et al. 2005; Toth and Robinson 2007; Carroll 2008;
105 Jovelin and Phillips 2009; Bergmiller et al. 2012), or under high pleiotropic
106 constraints due to either ubiquitous expression across tissues (McShea 2000; Mank et
107 al. 2008; Mank 2009; Meisel 2011) or due to high number of interactions with other
108 genes (i.e. high connectivity) (Provero et al. 2002; Carlson et al. 2006; Jasper et al.
109 2014), or displaying a large number of gene ontology (GO) terms (Choi et al. 2007;
110 Jovelin and Phillips 2009) are likely to be constrained in how readily they can be
111 expressed in a morph-biased manner, as their stable expression is necessary in both
112 morphs (Meisel 2011; Purandare et al. 2014).

113

114 Importantly, such constraints may hinder morph specific evolution, resulting in
115 antagonistic fitness variation, or genetic load. In other words, if selection pressures
116 are different in the different morphs for any given gene, selection cannot reach the
117 optimum for either caste, unless expression of the gene is caste specific (i.e. the gene
118 is downregulated in the morph where the its expression has negative effects). Such
119 constraints on the ability of the genome to respond adaptively to the divergent
120 selective pressures on the two morphs simultaneously have mainly been discussed
121 with respect to sexually antagonistic variation and sex biased gene expression (Rice
122 1998; Ellegren and Parsch 2007; Innocenti and Morrow 2010; Holman 2014) but the
123 same logic applies to castes in social insects, where "caste load" could potentially
124 ensue for genes whose caste specific expression evolution is constrained (Holman
125 2014). These constraints are the direct consequences of antagonistic selective

126 pressures for the caste phenotypes and genetic covariance between the reproductive
127 queens and the non-reproductive workers (Holman 2014).

128

129 In order to understand the constraints and flexibility of morph-biased expression
130 across species, and the potential for genetic load resulting from constraints on
131 expression evolution, comparisons of gene expression in several species that share the
132 same polyphenism are needed (Helanterä and Uller 2014). Although numbers of
133 species compared in studies to date are usually small, flexibility of morph-biased
134 expression patterns among species seems to be a general pattern, and only a very
135 small proportion of genes linked to conserved functions maintain a biased expression
136 pattern across species (Meiklejohn et al. 2003; Ellegren and Parsch 2007; Toth and
137 Robinson 2007; Toth et al. 2010; Fischman et al. 2011; Woodard et al. 2011; Wilkins
138 2013). This suggests that genes whose expression level and sequence are under
139 relatively weak purifying selection, are more likely to be expressed in a morph-biased
140 manner in any particular species, and ultimately even evolve in a morph-specific way
141 under positive selection (Helanterä and Uller 2014).

142

143 Caste differences in eusocial insects, such as ants, are one of the key model systems
144 for studying evolution of morph-biased gene expression (Gadau et al. 2011; Libbrecht
145 et al. 2013; Helanterä and Uller 2014). Originating from a common ancestor ca. 140–
146 168 Ma ago (Gadau et al. 2011), the reproductive division of labor between queen and
147 worker castes has facilitated enormous diversification especially in the worker castes
148 of ants, and dominance of many terrestrial ecosystems. In general, caste
149 differentiation is driven by early environmental differences, and queens and workers
150 arise from similar genetic backgrounds (Wilson 1971; Schwander et al. 2010).

151 Recently, a 16 species comparison of queen and worker transcriptomes covering the
152 major ant subfamilies showed that caste-biased expression patterns are highly variable
153 across species (Morandin et al. 2016). In this study, we investigate the determinants
154 of the variation in caste biases, and the potential for genetic load (“caste load”,
155 Holman 2014) resulting from constraints on caste-biased expression. We use three
156 measures of evolutionary constraint. The first one, average expression level, has been
157 shown to correlate with slow sequence evolution (Pál et al. 2001; Drummond et al.
158 2005a,b) and gene essentiality (Drummond et al. 2005b; Clark et al. 2007), i.e. the
159 severity of the harmful effects of gene deletion on fitness (Chen et al. 2012). Second,
160 connectivity of a gene in a co-expression network measures the number of
161 interactions a gene has with other genes (Provero et al. 2002). It is thus a correlate of
162 slow sequence evolution due to pleiotropic constraint, where the evolution of a
163 sequence is limited by its multiple roles. Third, the number of Gene Ontology (GO)
164 terms is similarly a measure of genes potentially having many biological functions,
165 and slow sequence evolution due to pleiotropic constraint.

166

167 More specifically, we will test six predictions. First, if the same features that correlate
168 with conserved sequence evolution also constrain caste-biased expression, we predict
169 that genes with high expression level, high connectivity and high number of GO terms
170 should have on average little caste-bias. Second, if these same features constrain caste
171 specific expression in any given species, we predict that they should be associated
172 with the correlation of expression levels among castes within species. Third, if these
173 same features constrain fast changes in expression patterns, the genes with high
174 expression level, high connectivity and high number of GO terms should show little
175 evolutionary flexibility in their expression patterns. Fourth, if constraints in the

176 evolution of gene expression patterns apply similarly to expression in both castes, and
177 thus produce “caste load”, we predict that expression variability across species should
178 show similar patterns in both queen and worker specific expression. Fifth, given that
179 worker phenotypes, and the selective pressures they face, have across the ant
180 phylogeny diverged more than queen phenotypes, we predict that caste specific
181 expression patterns should be more variable among species in workers. Finally, we
182 predict that genes showing little evolutionary flexibility in their expression patterns
183 would disproportionately include genes involved in conserved functions, such as
184 metabolism or development.

185

186 **Materials and Methods**

187 **Gene expression analysis.** In a previous study, we sequenced the transcriptomes of
188 queens and workers of 16 ant species from three subfamilies, which differed in a
189 variety of key traits (Figure 1) using Illumina Hi-Seq technology (Morandin et al.
190 2016). One species (*Lasius turcicus*) was not used in this study due to the lack of
191 queen samples. Three independent replicates for queens and workers, using whole-
192 body, were used in this study and variable number of samples were pooled per species
193 to obtain sufficient amount of RNA (Table S2). In total, we sequenced 100 libraries
194 from whole-body queen and worker samples, representing biological replicates of
195 each caste. In the absence of a reference genome for most species, we used a *de novo*
196 assembly methods to construct reference transcriptomes for each species separately
197 for use in read mapping for gene expression profiling using Trinity software (release
198 2013-02-25, default settings (Haas et al. 2013)). Paired-end reads were mapped to the
199 *de novo* transcriptomes using RSEM (Li and Dewey 2011) to obtain raw expression
200 values that were subsequently used in the Generalised Linear Models below.

201

202 **Weighted Gene Co-expression network to estimate Gene Connectivity.** Weighted

203 gene co-expression networks were constructed on the entire gene set using the mean

204 of normalized expression counts for each gene. A total of 9,859 gene expression data

205 were initially analyzed with the WGCNA package (Langfelder and Horvath 2008).

206 The input dataset consists of a table with each row representing one of the 9,859 gene

207 and each column one of the 100 samples (Table S2). Prior to the analysis, genes with

208 too many missing samples or zero variance were removed from the calculation as they

209 may affect our ability to detect gene co-expression. After this cleaning step, we

210 calculated the connectivity between each gene using the *softConnectivity* function

211 from the package WGCNA. More details on sample collection, cDNA synthesis,

212 library preparation, and weighted gene co-expression network analysis are described

213 by Morandin et al. (2016).

214

215 **Influence of constraints on the caste-biased expression patterns.** To test our first

216 hypothesis that genes with high expression level, high connectivity and high number

217 of GO terms are less likely to show caste-biased gene expression patterns on average,

218 we investigated the correlates of biased expression level using generalized linear

219 model (GLM) analysis (R Core Team 2012, version 3.1.1, glm function in the

220 package *stats*). First of all, in addition to the WGCNA cleaning steps described above,

221 genes with expression data from less than nine species were omitted from our study as

222 they would weaken our ability to detect a phylogenetic signal. Then we used each

223 gene as a sample ($n = 6608$), and quantified caste-bias for each gene with a non-

224 directional measure of the lower expression level (either queen or worker) divided by

225 the higher, using the average expression value across pools for each caste. This

226 produced a range of expression ratio values ranging from 0 to 1, where 0 indicates
227 extreme caste-bias with no expression in one caste and 1 represents equal expression
228 levels in the two castes. This value represents the average caste-bias in expression
229 level of each gene and was set as the response variable. As explanatory variables, we
230 used putative correlates of evolutionary constraints (obtained from Morandin et al.
231 2016), i.e. expression levels (a correlate of gene essentiality (Pál et al. 2003; Rocha
232 and Danchin 2004), shown to correlate with evolutionary rates in an earlier analysis,
233 Morandin et al. 2016), gene connectivity (a measure of pleiotropic constraint, shown
234 to correlate with evolutionary rates in Morandin et al. 2016) and the number of GO
235 terms (a potential correlate of pleiotropic constraint). All variables were log-
236 transformed to normalize the distribution of residuals before processing. In order to
237 find the best fitting model, we used the function *step* from the R software (R
238 development Core Team 2008) and the MASS package (Venables and Ripley 2002)
239 to choose a model using Akaike information criterion (AIC) with a stepwise
240 algorithm using the default values. Similar processes were used for all GLM analyses
241 below.

242

243 **Influence of constraints on the correlation in gene expression between queen and**
244 **worker.** Second, we tested the hypothesis that genes with high expression level, high
245 connectivity, and high numbers of GO terms would tend to be similar in their
246 expression among castes within species. The correlation of expression among castes
247 (similar to C_{mf} used in Dean and Mank 2016 for analyzing sex biased expression) for
248 each gene was calculated using the average queen and average worker gene
249 expression across all fifteen species. Only genes where expression levels were
250 available for both queen and worker samples within a species were used in this

251 analysis. The correlation coefficient for each gene was used as a response variable in
252 a GLM with gene connectivity, expression level and number of GO terms set as the
253 explanatory variables. Model selection was carried out as described above, and to
254 normalize the distribution of model residuals, the correlations of caste gene
255 expression were transformed prior to the analysis using the ‘powerTransformation’
256 function with Yeo-Johnson family in the ‘car’ library (Fox and Weisberg 2011) and a
257 subsequent Yeo-Johnson transformation using the ‘VGAM’ library (Yee 2016).

258

259 **Influence of constraints on the evolutionary flexibility of gene expression**

260 **patterns.** Third, we tested the hypothesis that genes with high expression level, high
261 connectivity, and high numbers of GO terms would vary little in their expression
262 values among closely related species. To do so, we investigated the influence of these
263 correlates of evolutionary constraints i.e. expression levels, gene connectivity and the
264 number of GO terms on the phylogenetic signal in expression levels using GLM
265 analysis. Phylogenetic signal is defined by Blomberg and Garland (2002) as the
266 “tendency of related species to resemble each other more than species drawn at
267 random from the same tree”. We used the Blomberg’s K values (Blomberg et al.
268 2003) to quantify the phylogenetic signal of expression levels across species with the
269 function *phylosig* from the R package *phytools* (Revell 2012), based on the phylogeny
270 in Morandin et al. (2016). The Blomberg’s test was chosen as it takes the
271 phylogenetic tree branch length into consideration in the analysis. Blomberg’s K
272 values close to 0 indicates phylogenetic independence, while high values indicate a
273 relationship between the phylogeny and the studied trait (Münkemüller et al. 2012).
274 Blomberg’s K values, calculated separately for expression level in queens and
275 workers, were set as the response variables in two separate GLM analyses, and gene

276 connectivity, number of GO terms, expression levels, and coefficient of variation in
277 expression among species were set as explanatory variables. The coefficient of
278 variation in expression level across species for each gene was added as an additional
279 explanatory variable to investigate the effect of overall variability on the phylogenetic
280 signal (Blomberg's K). For example, an extremely conserved gene could show zero
281 variation in its expression level, and consequently would not show any phylogenetic
282 signal either. The GLM analyses were conducted independently for expression levels
283 in queens and workers as the response variable. Transformations and model selection
284 were carried out as described above.

285

286 **Analysis of potential caste load and caste differences in expression variability.** If
287 the evolution of expression levels is similarly limited by general evolutionary
288 constraints in both castes, queens and workers should show correlated values of the
289 phylogenetic signals (Blomberg's K) and the overall levels of variation (measured as
290 coefficient of variation over species). We tested this prediction using Spearman's rank
291 correlation (R Core Team 2012, version 3.1.1. Because queens have a similar
292 reproductive function across species, while workers perform a comparatively more
293 diverse range of non-reproductive tasks, and face variable environments, in our
294 sample of species, we predicted that queen expression patterns would show stronger
295 phylogenetic signal than worker expression patterns (higher Blomberg's K values),
296 i.e. that caste specific expression patterns should be more variable among closely
297 related species in workers. Blomberg's K values for each gene (one value per gene
298 calculated using all species) were calculated separately using expression level in
299 queens and workers, and were compared between queen and worker expression
300 patterns using a paired t -test.

301 **Analysis of GO term enrichment for genes displaying a significant phylogenetic**
302 **signal.** Finally, we used GO term enrichment analysis to describe the functions of the
303 genes with significant phylogenetic signals ($p < 0.05$) to test the prediction that genes
304 with strong phylogenetic signals (high Blomberg's K values) have conserved
305 functions. Gene ontology (GO) terms for all genes were determined using
306 BLAST2GO (using BLASTp with an e-value cut-off $\leq 10^{-3}$) (Conesa et al. 2005).
307 Additionally, the GOstats package for R (Beissbarth and Speed 2004) was used to
308 conduct GO term enrichment analysis on genes presenting a significant phylogenetic
309 signal for queen or worker expression patterns ($p < 0.05$), as well as genes having a
310 phylogenetic signal common to both castes. We used the set of all genes for which
311 GO terms were available as the universe. Enriched GO terms ($FDR < 0.05$) were
312 subsequently clustered using Revigo (Supek et al. 2011) to avoid redundancy.

313

314 **Results**

315 **Influences of constraints on the caste-biased expression patterns.** After quality
316 filtering, 6608 genes comprising expression data from at least nine of the studied
317 species were retained for further analysis (mean number of species = 14, s.d. = 1.4).
318 The average caste biases for these genes are summarized in Figure S1, and the
319 frequency distribution of genes with significant expression bias for a given number of
320 species is given in Figure S2, showing that a majority of genes are never, or only very
321 rarely caste biased.

322

323 As predicted, genes with high average expression levels showed relatively little
324 average caste-bias (i.e. these genes had higher expression ratios, GLM, $p < 0.001$;
325 Table 1, Figure S3). However, and contrary to our initial predictions, neither

326 connectivity, nor the number of GO terms (Provero et al. 2002; Carlson et al. 2006;
327 Choi et al. 2007; Jovelin and Phillips 2009), had any significant relationship with the
328 ability of a gene to display a caste-biased expression patterns (Connectivity, GLM, p
329 = 0.65; Number of GO terms, GLM, $p = 0.07$, Table 1).

330

331 **Influence of constraints on the correlation of gene expression patterns among**
332 **castes.** According to our expectations, genes with high expression levels and also
333 genes that were highly connected showed a positive association with the correlation in
334 gene expression between queen and worker across the broad phylogenetic scale
335 (GLM, $p < 0.001$; Table 2 and Figure S4). However, there was also a negative
336 interaction among expression level and connectivity, such that the positive overall
337 effect of connectivity was the smaller, the higher the expression value of the gene was
338 (GLM, $p < 0.001$; Figure S4). The correlation between sequence evolutionary rates
339 (dN/dS , as calculated for each gene in Morandin et al 2016) and the among caste
340 correlation in expression was weakly negative ($p < 0.001$; $r_s = -0.11$; Figure S5,
341 supporting our interpretation that conserved genes are less likely to become caste
342 biased (Figure S2).

343

344 **Influence of constraints on evolutionary flexibility of gene expression patterns.**
345 Contrary to our expectations, genes with high expression levels had low Blomberg's
346 K values (e.g. low phylogenetic signals) (GLM, $p < 0.001$; Table 3, i.e. their
347 expression levels tended to be highly variable among closely related species).
348 However, highly connected genes showed high Blomberg's K values (i.e. low
349 variation among closely related species) for both castes as predicted (GLM, $p < 0.001$;
350 Table 3), especially if the genes also had high average expression (GLM, $p < 0.001$;

351 Table 3, Figure 2a and b). Finally, the coefficient of variation was negatively
352 correlated with the Blomberg's K values for both castes (GLM, $p < 0.001$; Table 3)
353 showing that high phylogenetic signals are not just artefacts of high variability
354 overall. The number of GO terms was not correlated with the Blomberg's K values in
355 either caste ($p = 0.50$, Table 3). Against our predictions, the Blomberg's K values
356 were not correlated with evolutionary rates (dN/dS) (Worker, $p = 0.08$; Queen, $p =$
357 0.77), however.

358

359 **Analysis of potential caste load and caste differences in expression variability.**

360 A total of 21 % of the genes showed a significant phylogenetic signal (Blomberg's K,
361 $p < 0.05$) for worker expression patterns (1379 genes), and 25 % for the queen
362 expression patterns (1636 genes). Among these genes, 754 genes (11 %) displayed a
363 significant phylogenetic signal for both queens and workers simultaneously, which is
364 significantly higher than expected if Blomberg's K values in queens and workers
365 were independent from each other (Fisher's test, $p < 0.001$). As predicted, queen and
366 worker Blomberg's K values were positively correlated (Spearman correlation, $r_s =$
367 0.49 , $n = 6608$, $p < 0.001$; Figure 3a), suggesting that for most genes, expression
368 patterns do not change independently between queens and workers in a species-
369 specific manner across the ant phylogeny. A similar pattern was also found for the
370 coefficient of variation (Spearman correlation, $r_s = 0.72$, $n = 6608$, $p < 0.001$; Figure
371 3b), suggesting that genes that are highly variable in expression levels across species
372 are likely to be variable for both queens and workers. As we further predicted, the
373 expression patterns in queens had stronger phylogenetic signals (higher Blomberg's K
374 values) than expression patterns in workers (mean_{worker} = 0.39, s.d._{worker} = 0.50, mean_{queen}
375 = 0.45, s.d._{queen} = 0.63, paired student t -test, $n = 6608$, $p < 0.001$, Figure S6,

376 Table S3), indicating that expression levels in queens are more likely to be similar
377 among closely related species than those in workers.

378

379 **Analysis of GO term enrichment for genes displaying a significant phylogenetic**
380 **signal.** Genes showing a significant phylogenetic signal in workers were enriched for
381 GO terms linked to “Mitotic DNA Damage Checkpoint” and “Metabolic Processes”.
382 For queens, terms such as “Metabolic Processes” and “Gene Expression Processes”
383 were enriched. Finally, genes that showed a phylogenetic signal in expression pattern
384 for both castes were enriched in terms related to “Development” and “Meiosis”. A
385 simplified figure of the GO terms and the complete list of enriched terms can be
386 found in the additional materials Figure S7 and Table S4.

387

388 **Discussion**

389 This study shows that not all genes are equally likely to be expressed in a morph-
390 biased manner in the queen-worker caste polyphenism of ants. Instead, features like
391 gene connectivity and average expression levels explain variation and average of
392 caste-biased expression, respectively, suggesting potential evolutionary constraints on
393 evolution of expression patterns. Patterns of variation in queen and worker
394 expression were correlated, suggesting that constraints of expression patterns in the
395 two castes are tightly associated in many genes, possibly leading to “caste load” that
396 limits independent evolution of the two morphs (Holman 2014). Our results highlight
397 the importance of analysing evolutionary constraints, such as those arising from gene
398 network context and conserved function, on the evolution of morph-biased gene
399 expression, and ultimately the decoupled evolution of polyphenic phenotypes.

400

401 **Influences of constraints on the caste-biased expression patterns.** The finding that
402 highly expressed genes are less likely to be caste-biased when averaged across species
403 fits our general argument about the shared factors constraining sequence and
404 expression evolution. Levels of expression are consistently negatively correlated with
405 rates of gene sequence evolution in multiple organisms, including the set of ant
406 species studied here (Krylov et al. 2003; Hahn and Kern 2005; Jovelin and Phillips
407 2009; Mikheyev and Linksvayer 2015; Morandin et al. 2016). Thus, evolutionary
408 constraints that play a role in sequence evolution are also likely to affect the evolution
409 of morph-biased expression patterns, and weakly constrained genes are likely to be
410 more easily co-opted for morph-specific use (Hunt et al. 2011; Leichty et al. 2012;
411 Helanterä and Uller 2014).

412

413 Against our predictions, gene connectivity did not significantly correlate with the
414 average tendency of a gene to be expressed in a caste-biased manner, even if
415 connectivity has been shown to correlate with pleiotropic interactions (Promislow
416 2004; Camps et al. 2007; Razeto-barry et al. 2011) and slow sequence evolution
417 (Hahn and Kern 2005; Jovelin and Phillips 2009; Mikheyev and Linksvayer 2015;
418 Morandin et al. 2016) in many studies. Since patterns of caste-bias are shared among
419 species for only a small minority of genes (this study, Berens et al. 2014; Feldmeyer
420 et al. 2014; Morandin et al. 2015, 2016), it seems that measuring average caste-bias
421 over species across the phylogeny is likely to conceal many of the interesting patterns.

422

423 Indeed, the correlation between connectivity and lack of caste bias was demonstrated
424 in the analysis where the correlation among castes within each species was
425 investigated. This suggests that highly connected genes, that are presumable highly

426 pleiotropically constrained, are less likely to vary among castes within species, but
427 their expression levels are coupled within species. This, together with positive
428 correlation among expression levels and the correlation among castes within species
429 suggests that understanding caste biases needs to take into account both the functional
430 context of the genes, and their variation across species (Morandin et al. 2016).

431

432 **Influence of constraints on evolutionary flexibility of gene expression patterns.**

433 Highly connected genes were more likely to have high phylogenetic signal in their
434 expression variation, i.e. tended to vary little among closely related species. The
435 observed positive statistical interactions suggest that this effect was particularly high
436 for genes with high average expression, even if gene expression level alone was
437 negatively correlated with the phylogenetic signals. In this context connectivity can
438 be seen as a measure of pleiotropic constraint (Hahn and Kern 2005; Kim et al. 2007);
439 for example, Carlson et al. (2006) found a relationship between gene connectivity and
440 sequence conservation in yeast. Also in previous studies looking at caste-biased gene
441 evolution in social insects, including earlier analyses of this same data set (Mikheyev
442 and Linksvayer 2015; Morandin et al. 2016), connectivity has been shown to be
443 negatively correlated with evolutionary rates. Similarly, high expression levels have
444 been shown to correlate with slow sequence evolution in this data set and others (Pál
445 et al. 2001; Drummond et al. 2005a; Morandin et al. 2016), and the correlation
446 between gene essentiality and high expression level in model organisms lends further
447 support to the use of expression level as a proxy for evolutionary constraint (Fraser et
448 al. 2004). Thus overall, these patterns suggest that factors such as connectivity, and to
449 a smaller effect also expression level, do affect both sequence evolutionary rates and
450 evolution of expression levels. For instance, stabilizing selection may maintain both

451 stable gene expression levels and sequence evolutionary rates across phylogeny
452 (Lemos et al. 2005; Laarits et al. 2016). However, causes and consequences are
453 difficult to tease apart, given the likely evolutionary (adaptive and neutral) feedbacks
454 between evolutionary rates, expression bias, average expression, and connectivity
455 (Helanterä and Uller 2014).

456

457 **Analysis of potential caste load and caste differences in expression variability.**

458 We found a strong positive correlation between phylogenetic signals and overall
459 variability observed in queen and worker gene expression patterns, suggesting that
460 genes with conserved expression pattern in both castes are unlikely to evolve caste
461 specific expression patterns. This result reinforces the concept of a “caste load”
462 suggested by Holman (2014), and the principle of morph antagonistic fitness variation
463 (Rice and Chippindale 2001; Morrow et al. 2008; Innocenti and Morrow 2010) and
464 antagonistic fitness variation in general (Immler et al. 2011). That is, genes whose
465 expression pattern is conserved in one caste, also typically retain a stable expression
466 pattern for both castes across the ant phylogeny due to e.g. pleiotropic constraint or
467 conserved function, impeding decoupling of caste phenotypes through differential
468 expression. We suggest that analysis of phylogenetic signals is a valuable tool for
469 understanding antagonistic fitness variation (see e.g. Dean and Mank 2016), and
470 should become more widely used as more and more phylogenetically informative
471 datasets become available. To date, antagonistic fitness variation has been thoroughly
472 assessed in relation to sex antagonistic variation and gender load, but we stress that it
473 is predicted to be a pervasive phenomenon applying to any polymorphism (Van
474 Dyken and Wade 2010; Snell-Rood et al. 2011; Holman 2014).

475

476 **Queen-worker differences in phylogenetic stability of expression pattern.** Lower
477 phylogenetic signal in gene expression patterns was found when looking at expression
478 patterns in workers, compared to queens. This result is following our expectations
479 because worker morphology, physiology and behavior have diverged more rapidly
480 than those of queens (e.g. evolution of several worker castes (Wheeler 1991)).
481 Moreover, queens have similar reproductive function across species and experience
482 less fluctuating environmental conditions due to the buffering provided by the nest
483 surroundings and the worker force. Furthermore, while workers undertake various
484 tasks within the colony, they also experience environmental changes faster than the
485 queens, which live in relatively well-buffered environments for most of their lives.
486 Thus, the higher variation in workers could reflect the effect of diverse environments
487 (e.g. *Formica* in diverse environments in the boreal zone, and *Monomorium chinense*
488 in the subtropical zone) on expression patterns being more pronounced in workers
489 than queens. Within a genus the environment may also dramatically vary (e.g.
490 *Formica aquilonia* in closed forest and *Formica cinerea* on open sand dunes).
491
492 Functional annotation of genes displaying a phylogenetic signal showed that gene sets
493 with conserved expression patterns were enriched in genes with metabolic and
494 developmental process functions, suggesting it is these genes that underlie large parts
495 of the possible caste load. Genes involved in these processes are also predominantly
496 under the control of purifying selection, showing low dN/dS ratios across species (e.g.
497 yeast (Hahn and Kern 2005; Vitkup et al. 2006; Lu et al. 2007); mammals
498 (Vinogradov 2015)). This finding is thus also in line with our assertion that factors
499 correlated to conservation at the sequence level are correlated with conservation at the
500 gene expression patterns as well. Our study is the first, to our knowledge, to study

501 conserved expression patterns in particular functional sets of genes across such a large
502 number of species. Metabolic and developmental processes include many conserved
503 functions making it possible that genes with such functions have universally relatively
504 stable expression patterns.

505

506 We used phylogenetic signal as a proxy of how likely a gene is to show variable
507 expression patterns among closely related species, thus reflecting the ease with which
508 expression patterns change. Given that a phylogenetic signal possibly reflects
509 outcomes of several different evolutionary processes, i.e. that variation in the
510 phylogenetic signal can emerge from variation in e.g. mutation rates and type or
511 strength of selection, (e.g. Revell et al. 2008; Kamilar et al. 2013), we have to remain
512 cautious in inferring the kind of evolutionary processes underlying variable
513 expression patterns, especially when we observed no correlation between
514 evolutionary rates and phylogenetic signals in expression levels. In other words, not
515 all genes showing low phylogenetic signal are necessarily ones where expression
516 patterns are evolving under strong positive or divergent selection, or unconstrained.
517 Such caution is especially important given the lack of a clear neutral baseline
518 expectation (comparable to neutral expectation in sequence evolution) for expression
519 evolution (Helanterä and Uller 2014), and our modest understanding of the prevalent
520 evolutionary processes underlying expression evolution (Khaitovich et al. 2006;
521 Warnefors and Kaessmann 2013; Rohlf et al. 2014; Hodgins-Davis et al. 2015),
522 including the role of neutral drift in expression evolution. Nevertheless, the observed
523 associations of connectivity, expression levels, total variation in expression levels,
524 and phylogenetic signals give us some confidence in interpreting low phylogenetic
525 signal as a measure of how flexible expression patterns of a given gene are.

526

527 Evolution of caste biased gene expression in social insects has interesting similarities
528 to evolution of tissue-specific gene expression. Indeed, if an insect society is seen as a
529 superorganism, queens and workers are analogous to reproductive and somatic
530 tissues, respectively (Helanterä 2016). As with social insect castes, it seems that genes
531 with a tissue-specific expression profile are prone to fast evolution due to low
532 pleiotropic constraint (Duret et al. 1999), but it remains an open question whether
533 genes with little constraint have been co-opted into tissue-specific expression pattern
534 or whether the evolutionary rates have increased after the expression patterns have
535 changed. Tissue specificity is also an important determinant of morph biased
536 expression, as genes with high specificity are more likely expressed in a sex biased
537 manner (Yang et al. 2006; Ellegren and Parsch 2007; Dean and Mank 2016).

538

539 **Conclusion.** Our study broadly suggests that constrained genes are less likely to be
540 involved in caste-biased expression. Conversely, genes evolving under lower
541 constraint are also likely to evolve morph-biased expression, and especially variable
542 morph-biased expression. Furthermore, relatively unconstrained genes are likely to
543 have expression patterns that are decoupled in the two castes, whereas constrained
544 genes are likely to share their expression patterns between castes. In other words,
545 genes that tend to operate under relatively little evolutionary constraints are more
546 likely to exhibit caste-biased expression patterns, and thus be co-opted for caste
547 specific use, in turn potentially leading to positive morph specific selection that may
548 resolve antagonistic selection pressures. Our results emphasize that large-scale
549 phylogenetic comparisons of morph-specific expression patterns and their correlates

550 are an essential tool in eventually understanding the roles of adaptive and non-
551 adaptive processes in polyphenic evolution.

552

553 **References**

554 Beissbarth, T., and T. P. Speed. 2004. GStat: find statistically overrepresented Gene
555 Ontologies within a group of genes. *Bioinformatics* 20:1464–1465.

556 Berens, A. J., J. H. Hunt, and A. L. Toth. 2014. Comparative transcriptomics of
557 convergent evolution: different genes but conserved pathways underlie caste
558 phenotypes across lineages of eusocial insects. *Mol. Biol. Evol.* 32:690–703.

559 Bergmiller, T., M. Ackermann, and O. K. Silander. 2012. Patterns of evolutionary
560 conservation of essential genes correlate with their compensability. *PLoS Genet.*
561 8:e1002803.

562 Blomberg, S. P., and T. Garland. 2002. Tempo and mode in evolution: Phylogenetic
563 inertia, adaptation and comparative methods. *J. Evol. Biol.* 15:899–910.

564 Blomberg, S. P., T. Garland, and A. R. Ives. 2003. Testing for phylogenetic signal in
565 comparative data: behavioral traits are more labile. *Evolution* 57:717–745.

566 Braendle, C., G. K. Davis, J. A. Brisson, and D. L. Stern. 2006. Wing dimorphism in
567 aphids. *Heredity.* 97:192–199.

568 Camps, M., A. Herman, E. Loh, L. A. Loeb, A. Herman, and E. Loh. 2007. Genetic
569 constraints on protein evolution. *Biochem. Mol. Biol.* 42:313–326.

570 Carlson, M. R. J., B. Zhang, Z. Fang, P. S. Mischel, S. Horvath, and S. F. Nelson.
571 2006. Gene connectivity, function, and sequence conservation: predictions from
572 modular yeast co-expression networks. *BMC Genomics* 7:40.

573 Carroll, S. B. 2008. Evo-Devo and an expanding evolutionary synthesis: a genetic
574 theory of morphological evolution. *Cell* 134:25–36.

575 Chen, W., P. Minguéz, M. J. Lercher, and P. Bork. 2012. OGEE : an online gene
576 essentiality database. *Nucleic Acids Res.* 40:901–906.

577 Choi, J. K., S. C. Kim, J. Seo, S. Kim, and J. Bhak. 2007. Impact of Transcriptional
578 Properties on Essentiality and Evolutionary Rate. *Genetics* 206:199–206.

579 Clark, A. G., M. B. Eisen, D. R. Smith, C. M. Bergman, B. Oliver, T. A. Markow, T.
580 C. Kaufman, M. Kellis, W. Gelbart, and V. N. Iyer. 2007. Evolution of genes and
581 genomes on the *Drosophila* phylogeny. *Nature* 450:203–218.

582 Conesa, A., S. Götz, J. M. García-Gómez, J. Terol, M. Talón, and M. Robles. 2005.
583 Blast2GO: a universal tool for annotation, visualization and analysis in functional
584 genomics research. *Bioinformatics* 21:3674–3676.

585 Dean, R., and J. E. Mank. 2016. Tissue specificity and sex-specific regulatory
586 variation permits the evolution of sex-biased gene expression. *Am. Nat.* 188:E74–
587 E84.

588 Devlin, R. H., and Y. Nagahama. 2002. Sex determination and sex differentiation in
589 fish: An overview of genetic, physiological, and environmental influences.
590 *Aquaculture* 208:191–364.

591 Drummond, D. A., J. D. Bloom, C. Adami, C. O. Wilke, and F. H. Arnold. 2005a.
592 Why highly expressed proteins evolve slowly. *Proc. Natl. Acad. Sci.* 102:14338–
593 14343.

594 Drummond, D. A., A. Raval, and C. O. Wilke. 2005b. A single determinant
595 dominates the rate of yeast protein evolution. *Mol. Biol. Evol.* 23:327–337.

596 Duret, L., D. Mouchiroud, and C. Bernard. 1999. Determinants of substitution rates in
597 mammalian genes : expression pattern affects selection intensity but not mutation rate.
598 *Mol. Biol. Evol.* 68–74.

599 Ellegren, H., and J. Parsch. 2007. The evolution of sex-biased genes and sex-biased

600 gene expression. *Nat. Rev. Genet.* 8:689–98.

601 Evans, J. D., and D. E. Wheeler. 1999. Differential gene expression between
602 developing queens and workers in the honey bee, *Apis mellifera*. *Proc. Natl. Acad.*
603 *Sci.* 96:5575–5580.

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

607 Ferreira, P. G., S. Patalano, R. Chauhan, R. Ffrench-Constant, T. Gabaldón, R. Guigó,
608 S. Sumner, T. Gabaldon, and R. Guigo. 2013. Transcriptome analyses of primitively
609 eusocial wasps reveal novel insights into the evolution of sociality and the origin of
610 alternative phenotypes. *Genome Biol.* 14:R20.

611 Fischman, B. J., S. H. Woodard, and G. E. Robinson. 2011. Molecular evolutionary
612 analyses of insect societies. *Proc. Natl. Acad. Sci.* 108:10847–10854.

613 Fox, J., and S. Weisberg. 2011. *An R Companion to Applied Regression*. Second
614 Edition, Sage Publications.

615 Fraser, H. B., A. E. Hirsh, G. Giaever, J. Kumm, and M. B. Eisen. 2004. Noise
616 minimization in eukaryotic gene expression. *PLoS Biol.* 2:834–838.

617 Gadau, J., M. Helmkampf, S. Nygaard, J. Roux, D. F. Simola, C. D. R. Smith, G.
618 Suen, and Y. Wurm. 2011. The genomic impact of 100 million years of social
619 evolution in seven ant species. *Trends Genet.* 28:14–21.

620 Guidugli, K. R., M.-D. Piulachs, X. Bellés, A. P. Lourenço, and Z. L. P. Simões.
621 2005. Vitellogenin expression in queen ovaries and in larvae of both sexes of *Apis*
622 *mellifera*. *Arch. Insect Biochem. Physiol.* 59:211–8.

623 Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M.
624 B. Couger, D. Eccles, B. Li, M. Lieber, M. D. Macmanes, M. Ott, J. Orvis, N. Pochet,

625 F. Strozzi, N. Weeks, R. Westerman, T. William, C. N. Dewey, R. Henschel, R. D.
626 Leduc, N. Friedman, and A. Regev. 2013. De novo transcript sequence reconstruction
627 from RNA-seq using the Trinity platform for reference generation and analysis. *Nat.*
628 *Protoc.* 8:1494–1512.

629 Hahn, M. W., and A. D. Kern. 2005. Comparative genomics of centrality and
630 essentiality in three eukaryotic protein-interaction networks. *Mol. Biol. Evol.* 22:803–
631 806.

632 Harpur, B. A., C. F. Kent, D. Molodtsova, J. M. D. Lebon, A. S. Alqarni, A. A.
633 Owayss, and A. Zayed. 2014. Population genomics of the honey bee reveals strong
634 signatures of positive selection on worker traits. *Proc. Natl. Acad. Sci.* 111:2614–
635 2619.

636 Helanterä, H. 2016. An organismal perspective on the evolution of insect societies.
637 *Front. Ecol. Evol.* 4:1–12.

638 Helanterä, H., and T. Uller. 2014. Neutral and adaptative explanations for an
639 association between caste-biased gene expression and rate of sequence evolution.
640 *Front. Genet.* 28:1381–1392.

641 Hodgins-Davis, A., D. P. Rice, J. P. Townsend, and J. Novembre. 2015. Gene
642 expression evolves under a house-of-cards model of stabilizing selection. *Mol. Biol.*
643 *Evol.* 32:2130–2140.

644 Holman, L. 2014. Caste load and the evolution of reproductive skew. *Am. Nat.*
645 183:84–95.

646 Hunt, B. G., L. Ometto, Y. Wurm, D. Shoemaker, S. Yi, L. Keller, and M.
647 Goodisman. 2011. Relaxed selection is a precursor to the evolution of phenotypic
648 plasticity. *Proc. Natl. Acad. Sci.* 108:15936–15941.

649 Hunt, B. G., S. Wyder, N. Elango, J. H. Werren, E. M. Zdobnov, S. Yi, and M.

650 Goodisman. 2010. Sociality is linked to rates of protein evolution in a highly social
651 insect. *Mol. Biol. Evol.* 27:497–500.

652 Immler, S., G. Arnqvist, and S. P. Otto. 2011. Ploidally antagonistic selection
653 maintains stable genetic polymorphism. *Evolution.* 66:55–65.

654 Innocenti, P., and E. H. Morrow. 2010. The sexually antagonistic genes of *Drosophila*
655 *melanogaster*. *PLoS Biol.* 8:e1000335.

656 Jasper, W. C., T. A. Linksvayer, J. Atallah, D. Friedman, J. C. Chiu, and B. R.
657 Johnson. 2014. Large-scale coding sequence change underlies the evolution of
658 postdevelopmental novelty in honey bees. *Mol. Biol. Evol.* 32:334–346.

659 Jovelin, R., and P. C. Phillips. 2009. Evolutionary rates and centrality in the yeast
660 gene regulatory network. *Genome Biol.* 10:R35.

661 Kamilar, J. M., N. Cooper, and P. T. R. S. B. 2013. Phylogenetic signal in primate
662 behaviour, ecology and life history. *Phil Trans R Soc B* 368:20120341.

663 Khaitovich, P., W. Enard, M. Lachmann, and S. Pääbo. 2006. Evolution of primate
664 gene expression. *Nat. Rev. Genet.* 7:693–702.

665 Kim, P. M., J. O. Korbel, and M. B. Gerstein. 2007. Positive selection at the protein
666 network periphery : Evaluation in terms of structural constraints and cellular context.
667 *Proc. Natl. Acad. Sci.* 104:20274–20279.

668 Krylov, D. M., Y. I. Wolf, I. B. Rogozin, and E. Koonin. 2003. Gene loss, protein
669 sequence divergence, gene dispensability, expression level, and interactivity are
670 correlated in eukaryotic evolution. *Genome Res.* 13:2229–2235.

671 Laarits, T., P. Bordalo, and B. Lemos. 2016. Genes under weaker stabilizing selection
672 increase network evolvability and rapid regulatory adaptation to an environmental
673 shift. *J. Evol. Biol.* 1–15.

674 Langfelder, P., and S. Horvath. 2008. WGCNA: an R package for weighted

675 correlation network analysis. BMC Bioinformatics 9:559.

676 Leichty, A. R., D. W. Pfennig, C. D. Jones, and K. S. Pfennig. 2012. Relaxed genetic
677 constraint is ancestral to the evolution of phenotypic plasticity. Integr. Comp. Biol.
678 52:16–30.

679 Lemos, B., B. R. Bettencourt, C. D. Meiklejohn, and D. L. Hartl. 2005. Evolution of
680 proteins and gene expression levels are coupled in *Drosophila* and are independently
681 associated with mRNA abundance, protein length, and number of protein-protein
682 interactions. Mol. Biol. Evol. 22:1345–54.

683 Li, B., and C. N. Dewey. 2011. RSEM: accurate transcript quantification from RNA-
684 Seq data with or without a reference genome. BMC Bioinformatics 12:323.

685 Libbrecht, R., P. R. Oxley, D. J. Kronauer, and L. Keller. 2013. Ant genomics sheds
686 light on the molecular regulation of social organization. Genome Biol. 14:212.

687 Lu, C., Z. Zhang, L. Leach, M. Kearsey, and Z. Luo. 2007. Impacts of yeast metabolic
688 network structure on enzyme evolution. Genome Biol. 8:407.

689 Mank, J. E. 2009. Sex chromosomes and the evolution of sexual dimorphism : lessons
690 from the genome. Am. Nat. 173:141–150.

691 Mank, J. E., and H. Ellegren. 2009. Are sex-biased genes more dispensable? Biol.
692 Lett. 5:409–412.

693 Mank, J. E., L. Hultin-Rosenberg, M. Zwahlen, and H. Ellegren. 2008. Pleiotropic
694 constraint hampers the resolution of sexual antagonism in vertebrate gene expression.
695 Am. Nat. 171:35–43.

696 Manousaki, T., A. Tsakogiannis, J. Lagnel, E. Sarropoulou, J. Z. Xiang, N.
697 Papandroulakis, C. C. Mylonas, and C. S. Tsigenopoulos. 2014. The sex-specific
698 transcriptome of the hermaphrodite sparid sharpsnout seabream (*Diplodus puntazzo*).
699 BMC Genomics 15:655.

700 McShea, D. W. 2000. Functional complexity in organisms : parts as proxies. Biol.
701 Philos. 15:641–668.

702 Meiklejohn, C. D., J. Parsch, J. M. Ranz, and D. L. Hartl. 2003. Rapid evolution of
703 male-biased gene expression in *Drosophila*. Proc. Natl. Acad. Sci. 100:9894–9.

704 Meisel, R. P. 2011. Towards a more nuanced understanding of the relationship
705 between sex-biased gene expression and rates of protein-coding sequence evolution.
706 Mol. Biol. Evol. 28:1893–900.

707 Mikheyev, A. S., and T. Linksvayer. 2015. Genes associated with ant social behavior
708 show distinct transcriptional and evolutionary patterns. Elife 4:e04775.

709 Morandin, C., K. Dhaygude, J. Paviala, K. Trontti, C. Wheat, and H. Helanterä. 2015.
710 Caste-biases in gene expression are specific to developmental stage in the ant
711 *Formica exsecta*. J. Evol. Biol. 28:1705–1718.

712 Morandin, C., M. M. Y. Tin, S. Abril, C. Gómez, L. Pontieri, M. Schiøtt, L.
713 Sundström, K. Tsuji, J. S. Pedersen, H. Helanterä, and A. S. Mikheyev. 2016.
714 Comparative transcriptomics reveals the conserved building blocks involved in
715 parallel evolution of diverse phenotypic traits in ants. Genome Biol. 1–19. Genome
716 Biology.

717 Morrow, E. H., A. D. Stewart, and W. R. Rice. 2008. Assessing the extent of genome-
718 wide intralocus sexual conflict via experimentally enforced gender-limited selection.
719 J. Evol. Biol. 21:1046–1054.

720 Münkemüller, T., S. Lavergne, B. Bzeznik, S. Dray, T. Jombart, K. Schiffers, and W.
721 Thuiller. 2012. How to measure and test phylogenetic signal. Methods Ecol. Evol.
722 3:743–756.

723 Nef, S., O. Schaad, N. R. Stallings, C. R. Cederroth, J. L. Pitetti, G. Schaer, S. Malki,
724 M. Dubois-Dauphin, B. Boizet-Bonhoure, P. Descombes, K. L. Parker, and J. D.

725 Vassalli. 2005. Gene expression during sex determination reveals a robust female
726 genetic program at the onset of ovarian development. *Dev. Biol.* 287:361–377.

727 Pál, C., B. Papp, and L. D. Hurst. 2003. Rate of evolution and gene dispensability.
728 *Nature* 421:496–497.

729 Pál, C., B. Papp, and L. D. Hurst. 2001. Highly expressed genes in yeast evolve
730 slowly. *Genetics* 158:927–931.

731 Parsch, J., and H. Ellegren. 2013. The evolutionary causes and consequences of sex-
732 biased gene expression. *Nat. Rev. Genet.* 14:83–7.

733 Pereboom, J. J. M., W. C. Jordan, S. Sumner, R. L. Hammond, and A. F. G. Bourke.
734 2005. Differential gene expression in queen-worker caste determination in bumble-
735 bees. *Proc. Biol. Sci.* 272:1145–52.

736 Perry, J. C., P. W. Harrison, and J. E. Mank. 2014. The ontogeny and evolution of
737 sex-biased gene expression in *Drosophila melanogaster*. *Mol. Biol. Evol.* 31:1206–
738 19.

739 Promislow, D. E. L. 2004. Protein networks, pleiotropy and the evolution of
740 senescence. *Proc. Biol. Sci.* 271:1225–34.

741 Provero, P., F. Teorica, and V. P. Giuria. 2002. Gene networks from DNA microarray
742 data : centrality and lethality. *arXiv Prepr. cond-mat/0207345* 2–7.

743 Purandare, S. R., R. D. Bickel, J. Jaquierey, C. Rispe, and J. A. Brisson. 2014.
744 Accelerated evolution of morph-biased genes in pea aphids. *Mol. Biol. Evol.*
745 31:2073–2083.

746 Razeto-barry, P., J. Diaz, D. Cotoras, and R. A. Vasquez. 2011. Molecular evolution,
747 mutation size and gene pleiotropy: a geometric reexamination. *Genetics*. 3: 877-885.

748 Revell, L. J. 2012. phytools: An R package for phylogenetic comparative biology
749 (and other things). *Methods Ecol. Evol.* 3:217–223.

750 Revell, L. J., L. J. Harmon, and D. C. Collar. 2008. Phylogenetic signal, evolutionary
751 process, and rate. *Syst. Biol.* 57:591–601.

752 Rice, W. R. 1998. Male fitness increases when females are eliminated from gene
753 pool : Implications for the Y chromosome. *Proc. Natl. Acad. Sci.* 95:6217–6221.

754 Rice, W. R., and A. K. Chippindale. 2001. Intersexual ontogenetic conflict. *J. Evol.*
755 *Biol.* 14:685–693.

756 Richards, S., Y. Liu, B. R. Bettencourt, P. Hradecky, S. Letovsky, R. Nielsen, K.
757 Thornton, M. J. Hubisz, R. Chen, R. P. Meisel, O. Couronne, S. Hua, M. A. Smith, P.
758 Zhang, J. Liu, H. J. Bussemaker, M. F. Van Batenburg, S. L. Howells, S. E. Scherer,
759 E. Sodergren, B. B. Matthews, M. A. Crosby, A. J. Schroeder, D. Ortiz-barrientos, C.
760 M. Rives, M. L. Metzker, D. M. Muzny, G. Scott, D. Steffen, D. A. Wheeler, K. C.
761 Worley, P. Havlak, K. J. Durbin, A. Egan, R. Gill, J. Hume, M. B. Morgan, G. Miner,
762 C. Hamilton, Y. Huang, L. Waldron, D. Verduzco, K. P. Clerc-blankenbunrg, I.
763 Dubchak, M. A. F. Noor, W. Anderson, K. P. White, A. G. Clark, S. W. Schaeffer, W.
764 Gelbart, G. M. Weinstock, and R. A. Gibbs. 2005. Comparative genome sequencing
765 of *Drosophila pseudoobscura*: Chromosomal , gene , and cis -element evolution.
766 *Genome Res.* 1–18.

767 Rocha, E. P. C., and A. Danchin. 2004. An analysis of determinants of amino acids
768 substitution rates in bacterial proteins. *Mol. Biol. Evol.* 21:108–116.

769 Rohlf, R. V., P. Harrigan, and R. Nielsen. 2014. Modeling gene expression evolution
770 with an extended ornstein-uhlenbeck process accounting for within-species variation.
771 *Mol. Biol. Evol.* 31:201–211.

772 Schrader, L., H. Helanter, and J. Oettler. 2016. Accelerated evolution of
773 developmentally biased genes in the tetraphenic ant *Cardiocondyla obscurior*. *Mol.*
774 *Biol. Evol.* msw240.

775 Schwander, T., N. Lo, M. Beekman, B. P. Oldroyd, and L. Keller. 2010. Nature
776 versus nurture in social insect caste differentiation. *Trends Ecol. Evol.* 25:275–82.

777 Snell-Rood, E. C., A. Cash, M. V Han, T. Kijimoto, J. Andrews, and A. P. Moczek.
778 2011. Developmental decoupling of alternative phenotypes: insights from the
779 transcriptomes of horn-polyphenic beetles. *Evolution* 65:231–45.

780 Sumner, S. 2014. The importance of genomic novelty in social evolution. *Mol. Ecol.*
781 23:26–28.

782 Sumner, S., J. J. M. Pereboom, and W. C. Jordan. 2006. Differential gene expression
783 and phenotypic plasticity in behavioural castes of the primitively eusocial wasp,
784 *Polistes canadensis*. *Proc. R. Soc. London. Ser. B, Biol. Sci.* 273:19–26.

785 Supek, F., M. Bošnjak, N. Škunca, and T. Šmuc. 2011. Revigo summarizes and
786 visualizes long lists of gene ontology terms. *PLoS One* 6:e21800.

787 Swain, A., and R. Lovell-badge. 1999. Mammalian sex determination : a molecular
788 drama Mammalian sex determination : a molecular drama. *Genes Dev.* 13:755–767.

789 Team, R. C. 2012. R: A language and environment for statistical computing. R
790 Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0.

791 Toth, A. L., and G. E. Robinson. 2007. Evo-devo and the evolution of social behavior.
792 *Trends Genet.* 23:334–341.

793 Toth, A. L., K. Varala, M. T. Henshaw, S. L. Rodriguez-Zas, M. E. Hudson, and G. E.
794 Robinson. 2010. Brain transcriptomic analysis in paper wasps identifies genes
795 associated with behaviour across social insect lineages. *Proc. Biol. Sci.* 277:2139–48.

796 Van Dyken, J. D., and M. J. Wade. 2010. The genetic signature of conditional
797 expression. *Genetics* 184:557–570. Genetics Soc America.

798 Venables, W. N., and B. D. Ripley. 2002. *Modern Applied Statistics with S*. Fourth
799 Edi. Springer, New York.

800 Vinogradov, A. E. 2015. Consolidation of slow or fast but not moderately evolving
801 genes at the level of pathways and processes. *Gene* 561:30–34. Elsevier B.V.

802 Vitkup, D., P. Kharchenko, and A. Wagner. 2006. Influence of metabolic network
803 structure and function on enzyme evolution. *Genome Biol.* 7:R39.

804 Warnefors, M., and H. Kaessmann. 2013. Evolution of the correlation between
805 expression divergence and protein divergence in mammals. *Genome Biol. Evol.*
806 5:1324–1335.

807 Wheeler, D. E. 1991. The developmental basis of worker caste polymorphism in ants.
808 *Am. Nat.* 138:1218.

809 Wilkins, A. S. 2013. “the genetic tool-kit”: the life-history of an important metaphor.
810 *Advances in evolutionary developmental biology.* John Wiley & Sons, Hoboken, NJ.

811 Wilson, E. O. 1971. *The insect societies.* Cambridge, Massachusetts, USA, Harvard
812 University Press.

813 Woodard, S. H., B. J. Fischman, A. Venkat, M. E. Hudson, K. Varala, S. A. Cameron,
814 A. G. Clark, and G. E. Robinson. 2011. Genes involved in convergent evolution of
815 eusociality in bees. *Proc. Natl. Acad. Sci. U. S. A.* 108:7472–7477.

816 Wurm, Y., P. Uva, F. Ricci, J. Wang, S. Jemielity, C. Iseli, L. Falquet, and L. Keller.
817 2009. Fourmidable: a database for ant genomics. *BMC Genomics* 10:5.

818 Yang, L., Z. Zhang, and S. He. 2016. Both Male-Biased and Female-Biased Genes
819 Evolve Faster in Fish Genomes. *Genome Biol.* 8:3433–3445.

820 Yang, X., E. E. Schadt, S. Wang, H. Wang, A. P. Arnold, L. Ingram-drake, T. A.
821 Drake, and A. J. Lusis. 2006. Tissue-specific expression and regulation of sexually
822 dimorphic genes in mice. *Genome Res.* 16:995–1004.

823 Yatsu, R., S. Miyagawa, S. Kohno, B. B. Parrott, K. Yamaguchi, Y. Ogino, H.
824 Miyakawa, R. H. Lowers, S. Shigenobu, L. J. Guillette Jr., and T. Iguchi. 2016. RNA-

825 seq analysis of the gonadal transcriptome during Alligator mississippiensis
826 temperature-dependent sex determination and differentiation. BMC Genomics 17.
827 BMC Genomics.
828 Yee, T. W. 2016. VGAM: Vector Generalized Linear and Additive Models. R
829 package version 1.0-2.

830

831

832 **Figures**

833 **Figure 1. Phylogenetic relationships of 15 ant species** (including two social forms
834 of *Solenopsis invicta*). The phylogenetic tree was constructed using gene alignments
835 with the software RAxML (v. 8), using 1427 genes and 3.59 Mb of sequence, and the
836 analysis was partitioned by gene and conducted under a GTRGAMMAI model
837 (Morandin et al. 2016). This tree was used to infer phylogenetic signals using the
838 Blomberg test.

839

840 **Figure 2.** Plots showing the relationship between the phylogenetic signals
841 (Blomberg's K values) and gene connectivity in relation to the average gene
842 expression for a) worker b) queen. Values were log-transformed.

843

844 **Figure 3. a) Relationship between phylogenetic signals (Blomberg's K values) for**
845 **queen and worker expression patterns.** Phylogenetic signals (i.e. the tendency of
846 related species to resemble each other more than species drawn at random from the
847 same tree (Blomberg and Garland 2002)) of female caste gene expression patterns are
848 significantly positively associated, and are thus unlikely to change independently in
849 queens and workers. Values were log-transformed. **b) Relationship between**
850 **coefficient of variation for queen and worker expression patterns.** Coefficient of

851 variation of female caste gene expression patterns are significantly positively
852 associated, and are thus unlikely to change independently from each other.

853

854 **Tables**

855 **Table 1. Influences of constraints on the caste-biased expression patterns.** Results
856 of the **GLM** analyses (see text for details) of the average caste bias in expression
857 (represented as the ratio of queen and worker expression level) across the 15 study
858 species.

859

860 **Table 2. Influence of constraints on correlation of caste gene expression.** Results
861 of the GLM analyses (see text for details) of the correlation of gene expression
862 between queens and workers across the 15 study species.

863

864 **Table 3. Influence of constraints on evolutionary flexibility of gene expression**
865 **patterns.** Results of the **GLM** analyses (see text for details) of the phylogenetic signal
866 (Blomberg's *K*) in expression patterns of queens and workers across the 15 study
867 species.

868

869

870 **Supplementary Materials**

871 **Figure S1. Frequency distribution of average caste-bias across species in**
872 **expression level of each gene.** We quantified caste-bias as expression ratio, i.e. the
873 lower expression level (either queen or worker) divided by the higher, using the
874 average expression value per caste. This produced a non-directional measure of
875 expression ratio with values ranging from 0 to 1, where 0 represents extreme caste-

876 bias with no expression in one caste and 1 represents equal expression levels in the
877 two castes.

878

879 **Figure S2. Barplot representing the number of genes found commonly**

880 **differentially expressed across species.** There are 2300 genes that are never found
881 differentially expressed (0 species) and only one gene found differentially expressed
882 across all 15 species (including two social forms of *Solenopsis invicta*).

883

884 **Figure S3. Boxplot depicting the distribution of caste-biased expression through**

885 **the genes expression levels.** The black line in the box represents the second quantile
886 (median) of caste biases and the upper and lower boundary of box means third
887 quantile and first quantile, respectively for a) Queen b) Worker.

888

889 **Figure S4. The relationship between gene connectivity and the correlation in**

890 **gene expression between queen and worker**, in relation to the average caste
891 expression levels. Values were log-transformed.

892

893 **Figure S5. The negative relationship between the rates of molecular evolution**

894 **(dN/dS) and the correlation in gene expression between queen and worker within**
895 **species.** Values were log-transformed.

896

897 **Figure S6. Barplot representing the mean Blomberg's K (phylogenetic signal)**

898 value for expression levels of queen and worker. Errors bars indicate the standard
899 errors.

900

901 **Figure S7. Plots of functional enrichment analysis.** Pie chart from enriched GO
902 terms in the set of genes showing a significant ($p < 0.05$) phylogenetic signal of gene
903 expression patterns in queen, worker, and both. Enrichment analysis was performed
904 using GOSTATS package for R (Beissbarth and Speed 2004), and enriched GO terms
905 were subsequently summarized to meaningful clusters for visualization using Revigo
906 (Supek et al. 2011).

907

908 **Table S1. Data frame used as input for the GLM analysis.** 6608 genes were
909 considered for the analysis. BLAST annotations and number of GO terms were
910 computed using Blast2GO (Conesa et al. 2005) and the BLASTp function. Gene
911 connectivity was calculated using WGCNA (Morandin et al. 2016). Phylogenetic
912 signals were computed using the Blomberg test.

913

914 **Table S2. Number of queens and workers used for RNA pooling before library**
915 **preparation.** Three replicates per species per caste were sequenced using an equal
916 number of samples in each of them. Number of pooled *Formica exsecta* samples can
917 be found in Morandin et al. (2015).

918

919 **Table S3. Numbers of genes with significant phylogenetic signal for queen and**
920 **worker expression patterns.** The phylogenetic signal was computed using the
921 Blomberg test.

922

923 **Table S4. GO terms.** List of enriched Gene Ontology annotation for genes with
924 significant ($p < 0.05$) phylogenetic signal observed for queen, worker expression
925 patterns and both (represented by the caste column). Phylogenetic signals were

926 computed using the Blomberg test. The GOstats package for R (Beissbarth and Speed
927 2004) was used to conduct GO term enrichment analysis.

928

929

930

931

932

933

934