Evidence for a conserved queen-worker genetic toolkit across slave-making ants and their ant hosts

B. Feldmeyer[1], C. Gstöttl[2], J. Wallner[2], E. Jongepier[3], A. Séguret[3], D. Grasso[4], E. Bornberg-Bauer[5], S. Foitzik[6], J. Heinze[2]

[1] Senckenberg Biodiversity and Climate Research Centre, 60325 Frankfurt am Main, Germany

[2] Zoology / Evolutionary Biology, University of Regensburg, 93040 Regensburg, Germany

[3]Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, 1098 XH Amsterdam, The Netherlands

[4]Department of Chemistry, Life Sciences and Environmental Sustainability, Università di Parma, 43124 Parma, Italy

[5]Institute for Evolution and Biodiversity, Westfälische Wilhelms University, 48149 Münster, Germany

[6]Institute of Molecular and Organismic Evolution, Johannes Gutenberg University, 55128 Mainz, Germany

Abstract

The ecological success of social Hymenoptera (ants, bees, wasps) depends on the division of labour between the queen and workers. Each caste is highly specialized in their respective function in morphology, behaviour and life history traits, such as lifespan and fecundity. Despite strong defences against alien intruders, insect societies are vulnerable to social parasites, such as workerless inquilines or slave-making (dulotic) ants. Here, we investigate whether gene expression varies in parallel ways between slave-making ants and their host ants across five independent origins of ant slavery in the *"Formicoxenus*-group" of the ant tribe Crematogastrini. As caste differences are often less pronounced in slave-making ants than non-parasitic ants, we also compare the transcriptomes of queens and workers in these species. We demonstrate a substantial overlap in expression differences between queens and workers across taxa, irrespective of lifestyle. Caste affects the transcriptomes much more profoundly than lifestyle, as indicated by 37 times more genes being linked to caste than to lifestyle and by multiple caste-associated gene modules with strong connectivity. However, several genes and one gene module are linked to the slave-making lifestyle across the independent origins, pointing to some evolutionary convergence. Finally, we do not find evidence for an interaction between caste and lifestyle, indicating that caste differences remain consistent even when species switch to a parasitic lifestyle. Our findings are a strong indication for the existence of a core set of genes whose expression is linked to the queen and worker caste in this ant taxon, supporting the "genetic toolkit" hypothesis.

Key words: Social parasitism, caste, transcriptomes, gene networks, slave-making ants, dulosis, selection

Introduction

The ecological success of social insects is based on the efficient division of labour between reproductives and non-reproductives, i.e., in the social Hymenoptera, the queens and workers (Wilson 1971; Hölldobler and Wilson 2009). Instead of producing their own offspring, workers help to raise the offspring of their mother or other related queens. The altruism of workers is explained by their relatedness to the recipients of their help, which is maintained by the closure of the society against unrelated freeloaders or parasites (Hamilton 1964, 1987). Nevertheless, several species have evolved sophisticated ways to infiltrate and usurp social insect colonies (Rabeling 2020), and among these are the charismatic "slave-making" or "dulotic ants" (Hölldobler and Wilson 1990; Buschinger 2009; Heinze and d'Ettorre 2001). Freshly mated, young queens of slave-making ants invade the nests of closely related, non-parasitic ant species, where they kill or expel the resident queen(s) and often also the adult workers. Host workers emerging from the conquered host brood take care of the slave-maker queen and her own offspring, maintain the nest, and forage for food. Slave-maker workers do not engage in normal worker chores (Hölldobler and Wilson 1990; Buschinger 2009) but instead raid neighbouring host nests and pillage their brood, thus replenishing or increasing the host force.

Slavery originated ten times independently in ants (Stoldt and Foitzik 2020) and at least five convergent origins lie within the "*Formicoxenus*-group" of the myrmicine tribe Crematogastrini (Blaimer et al. 2018). This allows investigation of whether convergent changes in gene expression occurred among those evolutionary switches to slave-making. There are several morphological, physiological and behavioural similarities among slave-making species of independent origin. For example, slave-maker workers are often heavily armed with strong mandibles and associated muscles, which leads to enlarged heads (Hölldobler and Wilson 1990). Given that workers of slave-making species no longer take care of the daily duties in the colony, they have become more queen-like in their task repertoire. They neither forage for food nor do they engage in brood care (Buschinger 2009). Moreover, slave-maker workers tend to have increased reproductive potential. While the ovaries of

non-parasitic workers have fewer ovarioles than the queen's ovaries and rarely contain mature eggs in the presence of a fertile queen, the ovaries of slave-maker workers often have the same number of ovarioles as the queen (Heinze 1996a). They form reproductive hierarchies (Franks and Scovell 1983; Bourke 1988; Heinze 1996b) and frequently lay male-destined eggs even in the queen's presence (Foitzik and Herbers 2001, Brunner et al. 2005, Suefuji and Heinze 2014). Gene expression differs strongly between workers and queens of most species, reflecting their divergent function, behaviour, and physiology (Gstöttl et al. 2020, Korb et al. 2021; Morandin et al. 2019a, b, Feldmeyer et al. 2014). For example, queen transcriptomes are characterized by the expression of genes associated with fecundity (e.g. vitellogenins), immunity, DNA repair and response to oxidative stress functionalities linked to their long lifespan (Stoldt et al. 2021). However, given the described similarities between slave-maker queens and workers, we expected their transcriptomes to differ less than those of queens and workers of non-parasitic species.

We therefore investigated the influence of caste (queen vs worker) and lifestyle (slave-maker vs non-parasitic) on gene expression of adult individuals and the interaction between those two parameters, which would indicate that caste is affecting gene expression differently in non-parasitic versus slave-making species. We used an experimental protocol that aimed for a large number of independent evolutionary origins of the slave-making lifestyle – five in total – all in the myrmicine tribe Crematogastrini. For each species, we sequenced the transcriptomes of six pooled workers and three pooled queens, respectively, taking slave-making species vs. host species as replicates. We investigated gene regulatory network properties according to caste and lifestyle and constructed orthologue clusters to investigate putative parallel selection patterns in genes associated to the slave-maker versus host lifestyle, including two related non-host taxa and two samples of the distantly related ant *Cardiocondyla obscurior* as outgroup.

Material & Methods

Sampling and sequencing

Colonies of the 15 myrmicine ant species of the "*Formicoxenus*-group" (genera *Harpagoxenus*, *Leptothorax*, and *Temnothorax*, including the previously synonymized genera of slave-making ants: *Chalepoxenus*, *Myrmoxenus*, and *Protomognathus*, Ward et al. 2015, but see Seifert et al. 2016) were collected between 2016-2018 from various locations across Germany, Italy, and the US (Supplement coordinates). Colonies were either brought to the lab in Regensburg, Mainz, or Münster, and kept under standard conditions in incubators (12 h 25°/ 12 h 25°C day-night cycles) before six workers and three queens were pooled per species for RNA extraction. For each species, workers and queens originated from three colonies. We generated one queen and one worker transcriptome for seven slave-making ants and their host species (*Harpagoxenus sublaevis / L. acervorum* and *L. muscorum,* Nürnberger Reichswald, Germany; *T. americanus (*formerly *Protomognathus americanus*) / *T. longispinosus*, New York, USA; *T. duloticus / T. curvispinosus,* Ohio, USA; *T. muellerianus* (formerly *Chalepoxenus muellerianus*) / *T. unifasciatus,* Calino nr. Rovato, Italy; *T. pilagens / T. ambiguus,* Michigan, USA; *T. ravouxi* (formerly *Myrmoxenus ravouxi*) / *T. unifasciatus*, Schönhofen, Germany; *T. mediterraneus* (formerly *Myrmoxenus kraussei*)*/ T. recedens,* Tignale, Italy). RNA was extracted using the Nucleo-Spin Mini kit (Macherey-Nagel). Samples were shipped to StarSEQ (Mainz) for library preparation and 100bp paired end sequencing on an Illumina HiSeq. In total, we obtained 15mio reads on average per sample (Supplement Table S2).

Gene expression analyses

Raw reads were quality checked using FastQC v.0.11.8 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and adapters trimmed with Trimmomatic v.2.8.4 (Bolger *et al.* 2014). HiSat2 v.2.1.0 (Kim et al. 2015) was used to map the reads to the *T. longispinosus* genome v.1 (GenBank accession: GCA_004794745.1_tlon_1.0; Kaur et al. 2019). We chose to use a single species genome as reference for all species to be able to later directly compare expression patterns between slave-making ants and their hosts as well as between queens and workers across species, with species as replicates. Non-host species were not included in this analysis. The counts table was created with HTSeq

(Anders et al. 2015). To prevent spurious results due to low read counts, we removed from the counts matrix genes with less than 10 reads in at least four samples before the subsequent differential gene expression analysis with DESeq2 (Love *et al.* 2014). We started with the full model ~Caste+Lifestyle+Caste:Lifestyle. The interaction turned out to be insignificant (no differentially expressed genes for the interaction), and we thus based all follow-up analyses on the two main factors only. All p-values were adjusted by false discovery rate (FDR) correction as implemented in DESeq2. To determine whether the significant slave-maker differentially expressed genes (DEGs) are more numerous than we would expect by chance, we ran 1000 permutations on the DEGs analysis in R. To gain a deeper understanding for the functionality of DEGs, we 1) conducted a functional enrichment analysis, 2) inferred pathways in which the DEGs are involved, and 3) used a word mining approach based on longevity and fecundity terms. In detail, we ran Interproscan v.5.39-77.0 (Jones et al. 2014) locally to obtain GO information using the *T. longispinosus* predicted proteome (GenBank accession: GCA_004794745.1_tlon_1.0; Kaur et al. 2019) as query. The GO enrichment analysis was performed with the R package TopGO (Alexa and Rahnenführer, 2016), using the 'parentchild' algorithm and the Fishers exact test for significance. Furthermore, the proteome was annotated with KEGG functional ortholog numbers using the BlastKOALA web utility (last accessed: 11.02.2021; Kanehisa et al. 2016) with 'Eukaryotes, Animals' specified as Taxonomy group. KEGG pathway affiliation of genes that were differentially expressed in each of the four comparisons (queens compared to workers and vice-versa; slave-makers compared to hosts and vice-versa) was assessed using the online utility of KEGG mapper with *Apis mellifera* as reference species (Kanehisa & Sato 2020), and visualized including the log2-fold change using pathview (Luo et al. 2013) in R v.3.6.3 (R Core team 2020). KEGG pathway enrichment was assessed with the enrichKEGG function implemented in clusterProfiler v.3.14.3 (Yu et al. 2012) where the universe was defined as the total set of *T. longispinosus* protein predictions with a KEGG functional ortholog annotation and organism 'ko'. For the text mining approach, we used gene annotations based on a BlastP search of the proteome versus the RefSeq invertebrate database, as query for a UniProt search. More specifically,

we extracted the gene function information for each gene with entries from *C. elegans, D. melanogaster, A. mellifera* and searched for terms related to fecundity and longevity (Supplement Table S3), both traits which we assumed to differ between queens and workers (script available from Negroni et al. 2021). We conducted an enrichment analysis by conducting a Fisher's exact test to test whether the number of genes with terms in fecundity or longevity within differentially expressed genes was higher than expected with respect to the complete proteome. The online tool Venny v.2.0.2 was used to generate the Venn diagram (https://bioinfogp.cnb.csic.es/tools/venny).

Selected genes analysis

For this analysis, we added transcriptomes of four taxa that are not known to be parasitized by slave-making ants, which acted as a biological control for comparisons of selection intensity between slave-makers and hosts: as outgroup, two populations of *Cardiocondyla obscurior* from Okinawa, Japan, and Ilhéus, Itabuna, and the two non-host species of the *Formicoxenus*group, *Temnothorax nylanderi* (Mainz, Germany) and *T. rugatulus* (Arizona, USA). We used Trinity v.2.8.6 (Grabherr et al 2011) with standard settings to construct species-specific transcriptomes using both worker and queen transcripts. Nucleotide sequences were translated into amino acid sequences with Transdecoder (https://github.com/TransDecoder). As *de novo* transcriptomes are known to contain many transcript fragments and well as isoforms, we constructed orthogroups across all species with OrthoFinder (Emms & Kelly 2015) including protein sequences derived from the *Temnothorax longispinosus* genome (Kaur et al 2019) to prevent the inflation of orthogroups. After orthogroup construction, the *T. longispinosus* protein sequences derived from the genome were removed from the orthogroups from all downstream summary statistics and analyses. Since we only obtained a handful of single copy orthologs, we used an inhouse script (Supplement_script) to retain only a single sequence per species based on the length and percent identity identified from the pair-wise blast results from OrthoFinder. We used Clustal-omega v.1.2.4 (Goujon et al. 2010) to construct sequence alignments for each orthogroup, which were trimmed with TrimAL v.1.4.1 (Capella-Gutierrez et al. 2009) and the following

169 settings: -gappyout -resoverlap 0.75 -seqoverlap 75 -backtrans. To test for signatures of positive
170 selection we used the codeml implementation in ete3 v.3.1.1 (Huerta-Cepas et al. 2016) running the
171 branch site test of selection as follows: ete3 evol --models bsA bsA1 --tests bsA, bsA1 --leaves --
172 internals. The cluster-specific tree topology as inferred by RAxML v.8.2.12 (Stamatakis 2014) was used
173 as input tree, and each species was coded as foreground branch consecutively. Finally, we blasted the
174 *T. longispinosus* sequence of each ortholog cluster with signature of selection versus the *T.*
175 *longispinosus* genome and only retained ortholog clusters with sequences mapping to a single location,
176 as further means of preventing putative paralogs. We further ran a GO enrichment analysis to test for
177 overrepresented functions among the selected genes (details above), and used the KEGG annotations
178 to investigate the pathways in which genes with signature of selection are involved.
179
180 *Network analysis*
181 To identify networks of co-expressed genes (modules), we constructed a weighted gene co-
182 expression network using the WGCNA (Langenfelder and Horwath 2008) package in R. We used all
183 genes which had passed the quality filtering step for the expression analysis (N = 8327), thus not only
184 the differentially expressed genes. Gene counts were normalized using the
185 *varianceStabilizingTransformation* function from DESeq2 (Love *et al.* 2014). Following the WGCNA
186 guidelines, we picked a soft-thresholding power of 8 for adjacency calculation. To associate modules
187 to either caste or lifestyle, we first calculated the modules' eigengene using the *moduleEigengenes*
188 function and tested for module trait correlation using the *corPvalueStudent* function. The hub gene of
189 each module (i.e., the gene with the highest connectivity within a module) was determined using the
190 *chooseTopHubInEachModule* function. Moreover, for modules associated with caste or lifestyle, we
191 tested whether the connectivity of genes with caste-specific expression and genes differed from the
192 connectivity of those which were not differentially expressed,. We ran linear models with connectivity
193 as response and caste-specificity (queen, worker, NA) as explanatory variable in R.
194

To determine the relevance of fecundity and longevity associated genes in lifestyle and caste associated modules, we downloaded a list of 123 genes which were assembled based on information from *Drosophila* and are part of the TI-J-LiFe pathways (TOR/IIS-JH-Lifespan and Fecundity) (Korb et al. 2021). We used the Flybase identifiers to download the corresponding *Drosophila* sequences and used a BlastX search to identify *T. longispinosus* proteins with a blast hit e-value < e-10. In cases where we had two or three hits, we selected the protein with the smallest e-value, with the longest match, which mostly corresponded to the lowest e-value. In cases where we had >3 blast hits, we created a sequence alignment using MACSE v.2.03 (Ranwez et al. 2011). A Maximum Likelihood phylogenetic tree with 1000 bootstrap replicates was constructed with RAxML with PROTGAMMAWAG as protein substitution model.

Results

We set out to investigate the effect of lifestyle and caste on gene expression patterns across 15 different species of ants contrasting slave-making ants versus their hosts, and queens versus workers. We were mainly interested in determining whether there are common toolsets of genes characteristic for a specific lifestyle or caste. We additionally tested for genes under selection in the above-mentioned slave-maker-host pairs plus four additional non-host outgroup species.

In general, gene expression patterns seem to be most similar amongst phylogenetically close species. The principal component analysis revealed a very strong phylogenetic effect with 71% of variance explained by PC1, which separated the *Harpagoxenus / Leptothorax* group from the *Temnothorax* species (Supplement Figure 1), while PC2, explaining 8% of the variance, separated queens from workers. Caste had a much stronger effect on gene expression compared to lifestyle, as also evident in the heatmap dendrogram, where samples cluster according to caste rather than lifestyle (Figure 1). As our analysis did not reveal an interaction between lifestyle and caste (0 DEGs associated to the interaction term), we present the gene expression plus gene network results according to the main effects, caste and lifestyle, and finally the results from the selection analysis.

221

222 *Lifestyle*

223 We detected 62 differentially expressed genes (DEGs) between lifestyles (host versus slave-maker
224 species) cumulative across the five origins of slavery (see supplement Table 1 DEGs). A permutation
225 test with 1000 iterations revealed that this is more than expected by chance (p = 0.04). Six genes were
226 consistently higher expressed in all seven slave-maker species compared to hosts, of which three were
227 annotated: "protein DVR-1 homolog", "sulfotransferase family cytosolic 1B member 1-like", and "ATP-
228 dependent DNA helicase II subunit 1-like". Only a single gene, "NADPH oxidase 5", showed higher
229 expression in all seven host species compared to slave-makers (Figure 2). Various metabolic processes
230 were significantly enriched among the differentially expressed genes between lifestyles (total = 21
231 processes). Among genes up-regulated in slave-makers, 14 processes were enriched, and three
232 functions were enriched in genes up-regulated in hosts, including "cell death" (Supplement_GO). In
233 addition to the functional level, we also investigated whether the genes in the lifestyle DEGs set were
234 overrepresented in specific pathways. There was no enrichment for the complete set of lifestyle DEGs,
235 nor in the separate host or slave-maker up-regulated gene-sets. This might be explained by the low
236 number of lifestyle-associated DEGs, and by the fact that only about half of these lifestyle-specific
237 genes could be KEGG annotated.

238 Host colonies are generally larger than slave-maker colonies, *i.e.*, host queens are more fecund
239 than slave-maker queens. Moreover, fecundity is often positively correlated with longevity in social
240 insects (Heinze and Korb 2016; Negroni et al. 2016). We therefore conducted a word mining approach
241 based on UniProt entries to investigate whether genes with known fecundity or longevity functionality
242 were overrepresented in the sets of DEGs between hosts and slave-makers (Table 1). We indeed
243 recovered more DEGs putatively associated with fecundity than expected by chance (N = 31; Fisher's
244 exact test, p-value = 6.619e-09), but not for the longevity associated genes (Table 1).

245 We obtained 10 modules of co-expressed genes (named after colours as given by the *WGCNA*
246 package), of which one module, red, was positively associated with the slave-making lifestyle (Figure

5). This module contained 191 genes, of which 19 were up-regulated in slave-makers and none in hosts. Genes associated with this lifestyle module were enriched for "reactive oxygen species metabolic processes", "oxidation reduction processes", or "fatty acid biosynthetic process" (Supplement WGCNA). Genes up-regulated in slave-makers had the highest connectivity in this module associated with lifestyle (Supplement Figure S2).

Caste

Caste had a much stronger effect on gene expression than lifestyle (χ^2= 2496.8, df = 1, p-value < 2.2e-16), with 2,321 differentially expressed genes (DEGs) between queens and workers. Around half of all caste-specific DEGs showed consistent higher expression in queens (N=738 out of 1,295) or in workers (N = 450 out of 1,026) across all species. For example, genes for an *insulin-like growth factor 2*, *mRNA-binding protein 1*, *maternal protein exuperantia*, *histone deacetylases*, and several *serine/threonine-protein kinases* were more highly expressed in queens of all 15 species, while *pro-corazonine like* had higher counts in workers of all but one species. The number of genes with consistent expression across taxa was significantly lower for the lifestyle comparison (χ^2= 36.865, df = 1, p-value = 1.266e-09).

As queens and workers strongly differ in fecundity and longevity, we used a word mining approach based on UniProt entries to investigate whether genes with known fecundity or longevity functionality were overrepresented in the sets of caste-specific DEGs (Table 1). Indeed, we detected more fecundity-associated terms (N = 1076; Fisher's exact test, p-value = 2.2e-16), and more longevity-associated terms (N = 412; Fisher's exact test, p-value < 0.0003) than expected by chance.

In the caste-specific DEG set, 39 gene ontology (GO) functions were significantly overrepresented, many of which were associated with metabolic and biosynthetic processes (Supplement_GO). Genes up-regulated in queens were enriched for 60 functions linked to various metabolic processes or stress responses. Among the worker up-regulated genes, 78 functions were significantly enriched, also belonging to metabolic and biosynthetic processes, but also oxidation-reduction. In addition to the functional enrichment, we conducted a KEGG-pathway enrichment

273 analysis. In total, we found 27 pathways enriched amongst the caste-specific DEGs, 12 for genes up-
274 regulated in queens, and 61 genes overexpressed in workers (Supplement_KEGG). Multiple putatively
275 reproduction-associated and repair pathways, such as "meiosis-yeast", "cell cycle", "DNA replication",
276 "RNA transport," or "ribosome biogenesis," were enriched in queens, and pathways such as "olfactory
277 transduction" and "longevity regulating pathway" (Figure 3) were enriched in genes up-regulated in
278 workers (Supplement_KEGG).

279 The gene network analysis resulted in five modules associated with caste, three of which were
280 positively associated with queens (yellow, green, magenta), and two with workers (black and blue)
281 (Figure 5). There was a trend for the pink module to also be associated with caste ($p = 0.07$), and it was
282 included in the following analyses (for details see: Supplement_WGCNA). In all three worker associated
283 modules (black, blue, pink), connectivity was highest for worker up-regulated genes, lowest for queen
284 up-regulated genes, and intermediated for genes that were not differentially expressed. The same
285 pattern was observed in queen-specific modules (yellow, green, magenta), in which the queen up-
286 regulated genes had the highest connectivity (Supplement_WGCNA).

287

288 *Comparison to other species*

289 We identified 84 genes associated to one of the 10 WGCNA modules that have also previously been
290 linked to fecundity or longevity in *Drosophila*. We obtained these from the "TI-J-LiFe" list containing
291 123 candidate genes (Korb et al. 2021). About half of these (N = 60) were found in the turquoise
292 module, which was neither associated to caste nor lifestyle, and eleven were found in the blue, caste-
293 linked module (Supplement_WGCNA). To determine whether overrepresented gene functions linked
294 to caste associated modules from our *Formicoxenus*-group data set could be extrapolated to other
295 ants or even termites, we compared our modules to Morandin et al. (2016) and Lin et al. (2021). Only
296 three functions were shared between enriched GO-terms of caste associated modules in our dulosis
297 data set (N = 46) and enriched GO-functions in *Formica* caste associated modules (N = 155; Morandin
298 et al. 2016), namely "cellular protein modification process", "protein modification process",

299 "monovalent inorganic cation transport" (Table 2). Eight terms were shared with caste associated
300 modules in termites (N = 277; Lin et al. 2021) (Table 2). No terms were shared amongst the three
301 studies. As queen and workers differ in fecundity and longevity, we investigated whether the same set
302 of fecundity-longevity associated genes of the "TI-J-LiFe" list (N = 84) which were found in the termites
303 (Lin et al. 2021) could be found in our data set . There were two genes, Ras64B and Kr-h1, associated
304 with a queen-worker co-expression module in termites and in Crematogastrini.
305
306 *Selected genes analysis*
307 The transcriptomes of the 19 taxa, including four samples of non-host species, consisted of 67,150-
308 155,629 transcripts with 32,584-51,743 open reading frames and 92-96% DOGMA completeness
309 (Supplement_transcriptomes). In total, we obtained 10,699 ortholog clusters of which 5,826 clusters
310 contained at least one transcript per species. 1,398 clusters remained after trimming and filtering for
311 single copies per species. We identified 660 signatures of selection across all species, which are
312 associated with 424 ortholog clusters, each mapping to a single location on the genome
313 (Supplement_codeml). Between 6-92 genes showed signatures of selection in each species
314 (Supplement_codeml), but there was no difference in the number of selected genes between lifestyles
315 (linear model, $F = 0.61$, d.f. = 2, $p = 0.56$). In slave-makers, genes with signatures of selection were
316 significantly enriched in functions such as protein modification, demethylation and energy
317 maintenance (Figure 4a, Supplement_codeml). In hosts and non-hosts, several regulatory and
318 metabolic functions are significantly overrepresented in genes with signatures of selection (Figure
319 4b+c).
320 We identified 114 genes that were both differentially expressed and additionally showed signatures
321 of selection (Supplement_overlap). Of these, a single gene was differentially expressed between
322 castes and lifestyles (TLON_06615-RA; pleckstrin homology domain-containing family F member 2
323 isoform X1), all other 113 were differentially expressed only between castes. 35 of these caste-
324 specific DEGs showed signatures of selection in hosts, 23 in non-hosts, and 57 in slave-makers.

Interestingly, many nucleotide metabolism pathways were included in the overlapping gene lists including purine, alanine, aspartate, valine, to name just a few, and the "mTOR signalling pathway" (Supplement_overlap).

Discussion

Eusocial insects are characterised by their sophisticated division of labour, which led to the evolution of different castes. Ant queens are the main reproductives, known for their long life of up to several decades and high fecundity (Keller and Genoud 1997), while the mostly infertile and short-lived workers take care of the brood, foraging, and nest defence (Hölldobler & Wilson 1990). In slave-making species, however, workers do not perform the "general" worker chores. They raid host nests in summer and often are permitted to lay male-destined eggs in the presence of the queen. Our study shows that whole-body transcriptomes of seven species pairs of the "*Formicoxenus* group" of Crematogastrini (Blaimer et al. 2018; *Harpagoxenus*, *Leptothorax,* and *Temnothorax*) differ considerably more with caste (queen vs. worker) than with lifestyle (slave-maker vs. non-parasitic species). Caste polyphenism is based on differential gene expression, and our transcriptome analyses show that the expression of 1,188 genes consistently differs between queens and workers of all studied species, regardless of lifestyle. Though the evolution of a parasitic lifestyle is associated with considerable changes in morphology and behaviour, only few transcriptomic shifts from host to slave-making species were consistent across the five independent origins of slave-making.

Approximately 40x more genes were differentially expressed between castes than between lifestyles (2321 vs. 62), and six out of ten co-expression modules were caste-associated in contrast to a single lifestyle module. This difference might in part reflect the single evolutionary origin of caste diphenism in ants, whereas slavery evolved repeatedly (Beibl et al. 2005; Feldmeyer et al. 2017; Prebus 2017). Nevertheless, as slave-making workers are often fertile and do not take over normal worker chores, we had expected to find a considerable interaction between caste and lifestyle in gene expression as well as in gene connectivity, as caste differences are less pronounced in slave-making

351 species. The lack of the interaction indicates that slave-maker and host queens and workers are rather
352 similar on a molecular level. Furthermore, it corroborates the result of a previous study where caste
353 differences also exceeded differences due to other traits, such as worker sterility, queen number, or
354 invasiveness (Morandin et al. 2016).

355 While in this study, half of the genes differentially expressed between queens and workers
356 show the same expression pattern across all species, another study identified only a single gene
357 similarly expressed between queens and workers a set of 16 species of multiple genera (Morandin et
358 al. 2016). The reason for this discrepancy may be explained by the different species relationships as
359 well as the underlying data basis. We studied species within a single, closely related clade and used
360 the genome of a single species as reference. In contrast, Morandin et al. (2016) investigated species
361 from five genera in different subfamilies based on de novo assembled transcriptomes. Additionally,
362 there is evidence for similarities in pathways across different lineages of social insects (ants, bees,
363 wasps) rather than a "common toolkit" of genes responsible for the caste phenotype (Berens et al.
364 2014). In our phylogenetically more restricted data set however, we identified 1,188 genes
365 representing the "core set" of queen-worker differences across the 15 species. For example, the gene
366 "*maternal protein exuperantia*", a maternal effect gene which is needed for proper localisation of the
367 *bicoid* RNA during oocye formation (de Olivereira et al 2017; McDonald et al. 1991), but also plays a
368 role in *Drosophila* spermatogenesis (Hazelrigg et al. 1990) was up-regulated in all queens. Also, "G1/S-
369 specific cyclin-D2" (cycD) was more strongly expressed in queens compared to workers in all but one
370 species. This gene is involved in cell cycle regulation and part of the "FoxO signalling" and "Wnt
371 signalling pathway", both important regulators of longevity. It could thus be associated with the
372 lifespan differences between these two castes. More generally, half of the genes differentially
373 expressed between caste were associated with the UniProt functionalities "fecundity" and "longevity."

374 Reflecting the convergent evolution of ant slavery (Beibl et al. 2005; Feldmeyer et al. 2017;
375 Prebus 2017), the different parasitic species not only show pronounced differences in their
376 morphology, but also in raiding behaviour (Brandt et al. 2006; Johnson 2008; Kleeberg and Foitzik

2016). Species-specific raiding patterns are mirrored by species-specific gene expression patterns (Alleman et al. 2018) and genes under selection (Feldmeyer et al. 2017). Gene expression differences between pairs of slave-makers and their hosts showed much more variation across species (representing different origins of slaver-making) than the differences between queens and workers, *i.e.*, there are many more idiosyncrasies in how slave-makers and hosts differ. Nevertheless, we found 62 genes that varied with lifestyle, including genes for fatty acid synthases. This could be indicative either of differences in fat synthesis and maybe storage between the two lifestyles, but these synthases may also be involved in the synthesis of cuticular hydrocarbons which are used as communication signals to discriminate species and castes (Leonhard et al. 2016).

As slave-makers are the derived lifestyle with species-specific behaviours and traits, we expected to find more genes under selection in slave-makers than in hosts and/or non-host species, however the number of selected genes did not different between lifestyles. Genes with signatures of selection were significantly enriched in functions such as protein modification, demethylation, and energy maintenance. An interesting candidate among the genes that showed signatures of selection in only one or a few of the species is venom protease-like in *T. muellerianus*. Though at present nothing is known on the composition of the venom used by *T. muellerianus* to kill host ants during slave-raids, many Hymenopteran venoms are proteases (Touchard et al. 2016). On a higher level, many metabolism pathways were included in this overlapping gene list, including purine, alanine, aspartate, valine, to name just a few, and the “mTOR signalling pathway”.

Conclusion

Social parasitism represents a derived state with clear differences to host and non-host species from morphology to behaviour. In most social parasites, queens and workers are more similar than in host species. Examining gene expression patterns and gene regulatory networks of 15 different ant species spanning five origins of slave-making, we expected to find an interaction between caste and lifestyle effects on gene expression patterns. Despite the phenotypic differences between

403 slavemaker and host castes, gene expression profiles are remarkably similar. We observed a very
404 strong and reliable effect of caste on gene expression. Within our species set, we were able to
405 identify a core-set of 1,188 caste-specific genes, which show a consistent expression pattern across
406 all species irrespective of lifestyle, pointing to a “genetic toolkit” in this set of related ant species.
407

408 **Acknowledgements**

409 We would like to thank Tilman Shell for writing the ortholog cluster filter script for us. This work was
410 supported by the Deutsche Forschungsgemeinschaft (grant numbers Bo 2544/12-1 to EBB, Fo 298/20
411 to FO, He 1623/40 to JH).
412

References

Alexa A, Rahnenführer J (2016) Gene set enrichment analysis with topGO. R package version 2.24. 0

Anders S, Pyl PT, Huber W (2015) HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31: 166-169

Beibl J, Stuart RJ, Heinze J, Foitzik S (2005) Six origins of slavery in formicoxenine ants. Insect Soc 52: 291–297

Berens AJ, Hunt JH, Toth AL (2014) Comparative transcriptomics of convergent evolution: different genes but conserved pathways underlie caste phenotypes across lineages of eusocial insects. Mol Biol Evol 32: 690–703.

Blaimer BB, Ward PS, Schultz TR, Fisher BL, Brady SG (2018) Paleotropical diversification dominates the evolution of the hyperdiverse ant tribe Crematogastrini (Hymenoptera: Formicidae). Insect Syst Div 2: 1-14

Bolger AM, Lohse M, Usadel (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114-2120

Bourke AFG (1988) Dominance orders, worker reproduction, and queen-worker conflict in the slave-making ant *Harpagoxenus sublaevis*. Bheav Ecol Sociobiol 23: 323-333

Brandt, M., Heinze, J., Schmitt, T. & Foitzik, S. 2006.Convergent evolution of the Dufour's gland secretion as a propaganda substance in the slave-making ant genera *Protomognathus* and *Harpagoxenus*. Insectes Soc 53: 291–299

Brunner Em Trindl A, Falk KH, Heinze J, d'Ettorre P (2005) Reproductive conflict in social insects: male production by workers in a slave-making ant. Evolution 59: 2480-2482

Buschinger A (2009) Social parasitism among ants: a review (Hymenoptera: Formicidae). Myrmecol News 12: 219-235

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25: 1972-1973

de Oliveira, J.L., Sobrinho-Junior, I.S., Chahad-Ehlers, Alves de Brito R (2017) Evolutionary coincidence of adaptive changes in exuperantia and the emergence of bicoid in Cyclorrhapha (Diptera). Dev Genes Evol 227, 355–365. https://doi.org/10.1007/s00427-017-0594-3

Emms DM, Kelly S (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol 16: 157

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

Feldmeyer B, Elsner D, Alleman A, Foitzik S (2017) Species-specific genes under selection characterize the co-evolution of slavemaker and host lifestyles. BMC Evol Biol 17: 237

Foitzik S, Herbers JM (2001) Colony structure of a slavemaking ant. I. Intracolony relatedness, worker reproduction, and polydomy. Evolution 55: 307-315

Franks NR, Scovell E (1983) Dominance and reproductive success among slave-making worker ants. Nature 304: 724-725

Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R (2010) A new bioinformatics analysis tools framework at EMBL-EBI. Nucleic Acids Res 38: W695–W699.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nature Biotechnol 29: 644-652

Gstöttl C, Stoldt M, Jongepier E, Bornberg-Bauer E, Feldmeyer B, Heinze J, Foitzik S (2020) Comparative analyses of caste, sex, and developmental stage-specific transcriptomes in two Temnothorax ants. Ecol Evol 10: 4193-4203

Hamilton WD (1964) The genetical evolution of social behaviour, I.II. J theor Biol 52, 1-52

Hamilton WD (1987) Kinship, recognition, disease, and intelligence: constraints of social evolution. In: Ito Y, Brown JL, Kikkawa J (Eds.), Animal Societies: Theories and Facts, Japan Scientific Societies Press, Tokyo, pp. 81-102

Hazelrigg, T., Watkins, W. S., Marcey, D., Tu, C., Karow, M., & Lin, X. R. (1990). The exuperantia gene is required for Drosophila spermatogenesis as well as anteroposterior polarity of the developing oocyte, and encodes overlapping sex-specific transcripts. Genetics, 126(3), 607–617. https://doi.org/10.1093/genetics/126.3.607

Heinze J (1996a) The reproductive potential of workers in slave-making ants. Insectes Soc 43: 319-328

Heinze J (1996b) Reproductive hierarchies among workers of the slave-making ant, *Chalepoxenus muellerianus*. Ethology 102: 117-127

Heinze J, d'Ettorre P (2001) Sociobiology of slave-making ants. Acta Ethol 3: 67–82

Hölldobler B, Wilson EO (1990) The Ants. Harvard University Press, Cambridge, Mass.

Hölldobler B, Wilson EO (2009) The superorganism: the beauty, elegance, and strangeness of insect societies. WW Norton, New York

Huerta-Cepas J, Serra F, Bork P (2016) ETE 3: Reconstruction, analysis and visualization of phylogenomic data. Mol Biol Evol 33: 1635-1638 doi: 10.1093/molbev/msw046

Johnson CA (2008) Slave-maker ant competition for a shared host and the effect on coevolutionary dynamics. Ecol Monogr 78: 445–460

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics 30: 1236-1240

Jongepier E, Kleeberg I, Foitzik S (2015) The ecological success of a social parasite increases with manipulation of collective host behavior. J evol Biol 28: 2152–2162

Kanehisa, M. and Sato, Y. (2020) KEGG Mapper for inferring cellular functions from protein sequences. Protein Sci. 29, 28-35.

Kanehisa, M., Sato, Y., and Morishima, K. (2016) BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J. Mol. Biol. 428, 726-731.

Kaur R, Stoldt M, Jongepier E, Feldmeyer B, Menzel F, Bornberg-Bauer E, Foitzik S (2019) Ant behaviour and brain gene expression of defending hosts depend on the ecological success of the intruding social parasite. Phil Trans R Soc B 374: 20180192

Keller L, Genoud M (1997) Extraordinary lifespans in ants: a test of evolutionary theories of ageing. Nature 389: 958-960

Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. Nature Methods 12: 357-360

Kleeberg I, Foitzik S (The placid slavemaker: avoiding detection and conflict as an alternative, peaceful raiding strategy. Behav Ecol Sociobiol 70: 27–39

Korb J, Meusemann K, Aumer D, Bernadou A, Elsner D, Feldmeyer B, Foitzik S, Heinze J, Libbrecht R, Lin S, Majoe M, Monroy Kuhn JM, Nehring V, Negroni MA, Paxton RJ, Séguret AC, Stoldt M, Flatt T (2021) Comparative transcriptomic analysis of the mechanisms underpinning ageing and fecundity in social insects. Phil Trans R Soc B 376: 20190728

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

Sara Diana Leonhardt, Florian Menzel, Volker Nehring, Thomas Schmitt (2016) Ecology and Evolution of Communication in Social Insects. Cell Volume 164, Issue 6,Pages 1277-1287, doi.org/10.1016/j.cell.2016.01.035.

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15: 550

Luo, W. and Brouwer C., Pathview: an R/Bioconductor package for pathway-based data integration and visualization. Bioinformatics, 2013, 29(14): 1830-1831, doi: 10.1093/bioinformatics/btt285

MacDonald PM, Luk SK, Kilpatrick M (1991) Protein encoded by the exuperantia gene is concentrated at sites of bicoid mRNA accumulation in Drosophila nurse cells but not in oocytes or embryos. Genes & Development 5: 2455-2466. doi:10.1101/gad.5.12b.2455

Morandin C, Tin MMY, Abril S, Gómez C, Pontieri L, Schiøtt M, Sundström L, Tsuji K, Pedersen JS, Helanterä H, Mikheyev AS (2016) Comparative transcriptomics reveals the conserved building blocks involved in parallel evolution of diverse phenotypic traits in ants. Genome Biol 17: 43 https://doi.org/10.1186/s13059-016-0902-7

Morandin C, Hietala A, Helanterä H (2019a) Vitellogenin and vitellogenin-like gene expression patterns in relation to caste and task in the ant *Formica fusca*. Insectes Soc 66: 519-531

Morandin C, Brendel VP, Sundström L, Helanterä H, Mikheyev AS (2019b) Changes in gene DNA methylation and expression networks accompany caste specialization and age-related physiological changes in a social insect. Mol Ecol 28: 1975–1993

Negroni M.A., Macit M., Stoldt M., Feldmeyer B., S. Foitzik (2021) Molecular regulation of lifespan extension in fertile ant workers. Philosophical Transactions B 376:20190736 https://doi.org/10.1098/rstb.2019.0736

Prebus M (2017) Insights into the evolution, biogeography and natural history of the acorn ants, genus *Temnothorax* Mayr (Hymenoptera: Formicidae). BMC Evol Biol 17: 250 https://doi.org/10.1186/s12862-017-1095-8

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rabeling C (2020) Social parasitism. In: Starr C, Ed.) Encyclopedia of Social Insects, Springer, Cham. https://doi.org/10.1007/978-3-319-90306-4_175-1

Ranwez V, Harispe S, Delsuc F, Douzery EJP (2011) MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. PLoS One 6: e22594

536 Seifert B, Kleeberg I, Feldmeyer B, Pamminger T, Jongepier E, Foitzik S (2014) *Temnothorax pilagens*
537 sp. n. – a new slave-making species of the tribe Formicoxenini from North America (Hymenoptera,
538 Formicidae). ZooKeys 368: 65-77

539 Seifert B, Buschinger A, Aldawood A, Antonova A, Bharti H, Borowiec L, Dekoninck W, Dubovikoff D,
540 Espadaler X, Flegr J, Georgiadis C, Heinze J, Neumeyer R, Ødegaard F, Oettler J, Radchenko A, Schultz
541 R, Sharaf M, Trager J, Vesnić A, Wiezik M, Zettel H (2016) Banning paraphylies and executing
542 Linnaean taxonomy is discordant and reduces the evolutionary and semantic information content of
543 biological nomenclature. Insectes Soc 63: 237-242

544 Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
545 phylogenies. Bioinformatics 30: 1312-1313

546 Stoldt M, Foitzik S (2020) Slave-Making in Ants (Dulosis). In: Starr C. (Ed.) Encyclopedia of Social
547 Insects. Springer, Cham. https://doi.org/10.1007/978-3-319-90306-4_105-1

548 Stoldt M, Klein L, Beros S, Butter F, Jongepier E, Feldmeyer B, Foitzik S (2021) Parasite presence
549 induces gene expression changes in an ant host related to immunity and longevity. Genes 12: 95

550 Suefuji M, Heinze J (2014) The genetic population structure of two socially parasitic ants: ‘‘the active
551 slave-maker’’ *Myrmoxenus ravouxi* and the ‘‘degenerate slave-maker’’ *M. kraussei.* Conserv Genet
552 15: 201–211

553 Touchard A, Aili SR, Paterson Fox EG, Escoubas P, Orival J, Nicholson GM, Dejean A (2016) The
554 biochemical toxin arsenal from ant venoms. Toxins 8: 30

555 Ward PS, Brady SG, Fisher BL, Schultz TR (2015) The evolution of myrmicine ants: phylogeny and
556 biogeography of a hyperdiverse ant clade (Hymenoptera: Formicidae). Syst Entomol 40: 61-81

557 Wilson EO (1971) The Insect Societies. Harvard University Press, Cambridge, Mass.

558 Yu G, Wang L-G, Han Y, He Q-Y (2012) clusterProfiler: an R Package for Comparing Biological Themes
559 Among Gene Clusters. Omics J Integr Biol 16: 284-287

560

561 **Data Accessibility**

562 All raw sequence data underlying this study have been deposited in the National Centre for
563 Biotechnological Information (NCBI) Sequence Read Archive (SRA) and will be accessible upon
564 publication of this manuscript (BioProject accession number XY will follow in the next version of the
565 manuscript).

566 **Author Contributions**

567 The study was conceived by JH, EBB and SF, and was designed by EJ, BF, JH, EBB and SF. DG provided
568 ant samples. BF, CS, JW and EJ conducted the analyses. All authors contributed to writing the paper.
569 Authors declare no conflict of interest.

Table 1: Number of differentially expressed genes with a fecundity or longevity functionality based on a text mining approach using the UniProt database as a reference. Results of Fisher's exact test are given indicating whether the number of genes with either fecundity or longevity association in caste and lifestyle are more than expected by chance.

Treatment	**Fecundity**		**Longevity**	
	N of genes	**P-Value**	**N of genes**	**P-value**
Caste	**1076**	**2.2e-16**	**412**	**0.0003**
Queen	**630**		**262**	
Worker	**447**		**150**	
Lifestyle	**31**	**6.619e-09**	**13**	**0.39**
Host	**9**		**6**	
Slave-maker	**23**		**7**	

Table 2: Shared enriched GO terms between queen-worker caste associated modules from this study, a study on 16 ant species from multiple genera (Morandin et al. 2016), and the termite *Cryptotermes secundus* (Lin et al. 2021).

Organism(s)	Shared enriched GO-terms with this study
Multiple ants	cellular protein modification process
Multiple ants	protein modification process
Multiple ants	monovalent inorganic cation transport
C. secundus	cyclic nucleotide biosynthetic process
C. secundus	signal transduction
C. secundus	transmembrane transport
C. secundus	DNA replication initiation
C. secundus	transcription
C. secundus	G protein-coupled receptor signaling pathway
C. secundus	intracellular protein transport
C. secundus	proteolysis

Figure 1: Sample dendrogram and heatmap depicting a strong clustering according to caste and a weaker effect of lifestyle. The heatmap is based on the top 50 highest expressed genes. Slave-making species are highlighted in bold.

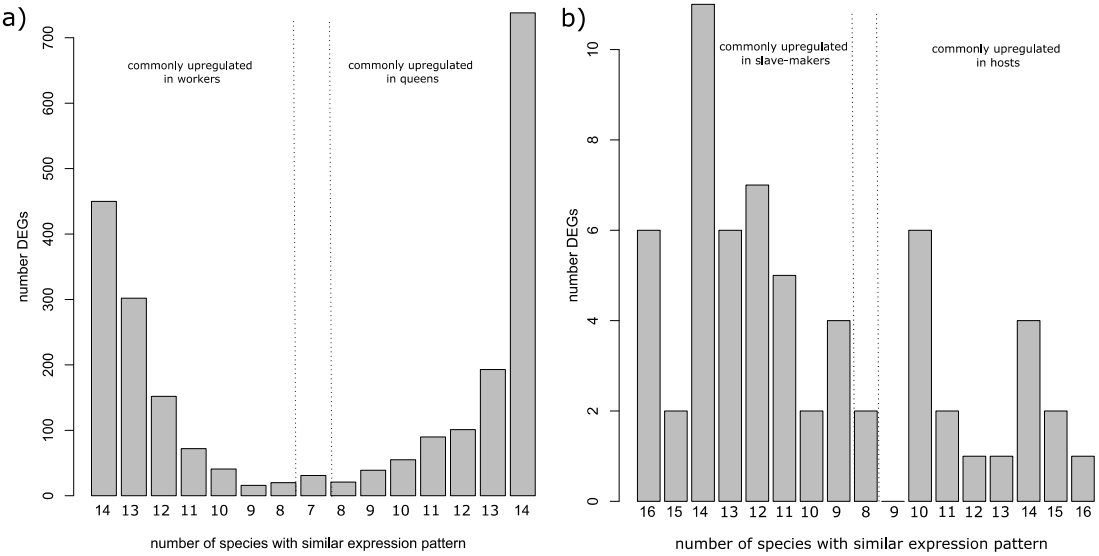

Figure 2: Number of species with similar gene expression patterns in a) queen and worker, or b) host and slave-maker comparisons. Please note that the host-slave-maker pairs add up to 16 comparisons instead of 14, since *Harpagoxenus sublaevis* parasitizes two host species *L. acervorum* and *L. muscorum.*

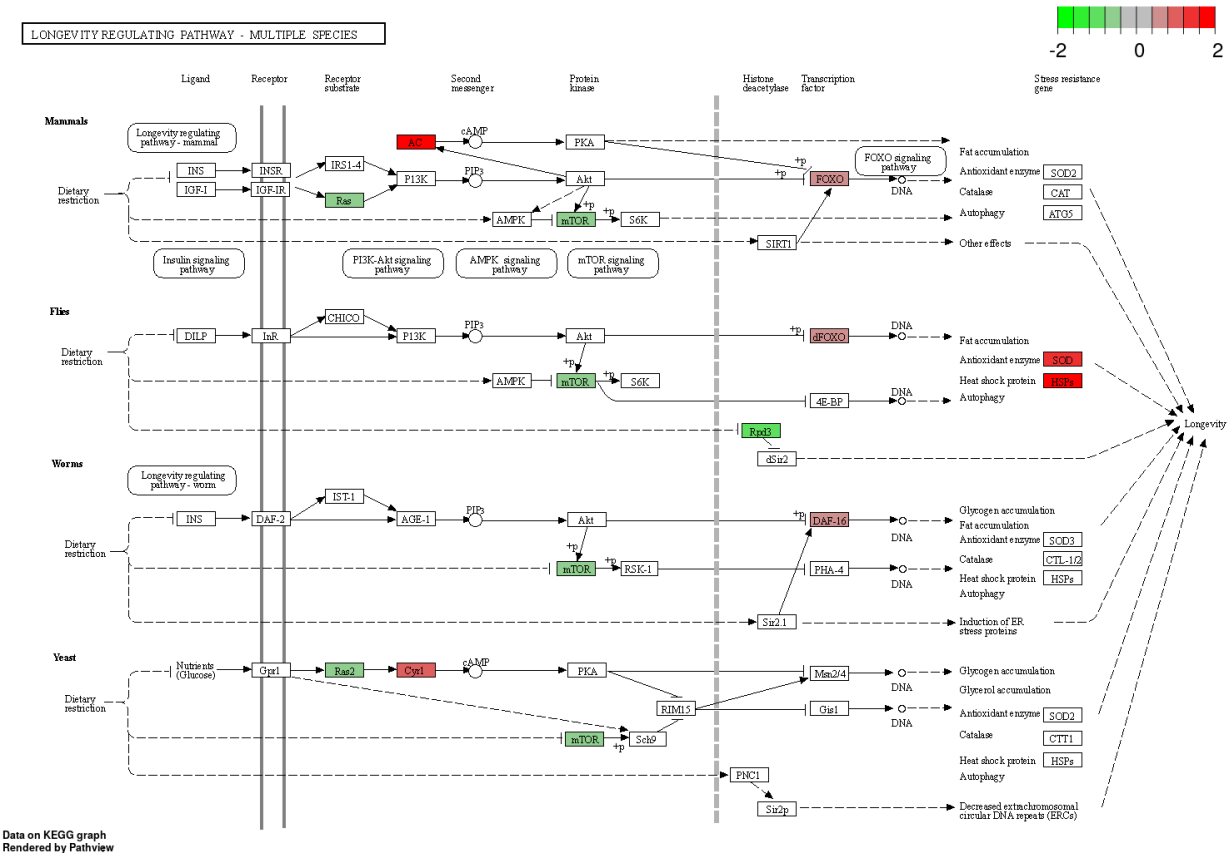

Figure 3: KEGG map depicting the "longevity regulating pathway" with up-regulated genes in queens (negative logFC; green) versus up-regulated in workers (positive logFC; red).

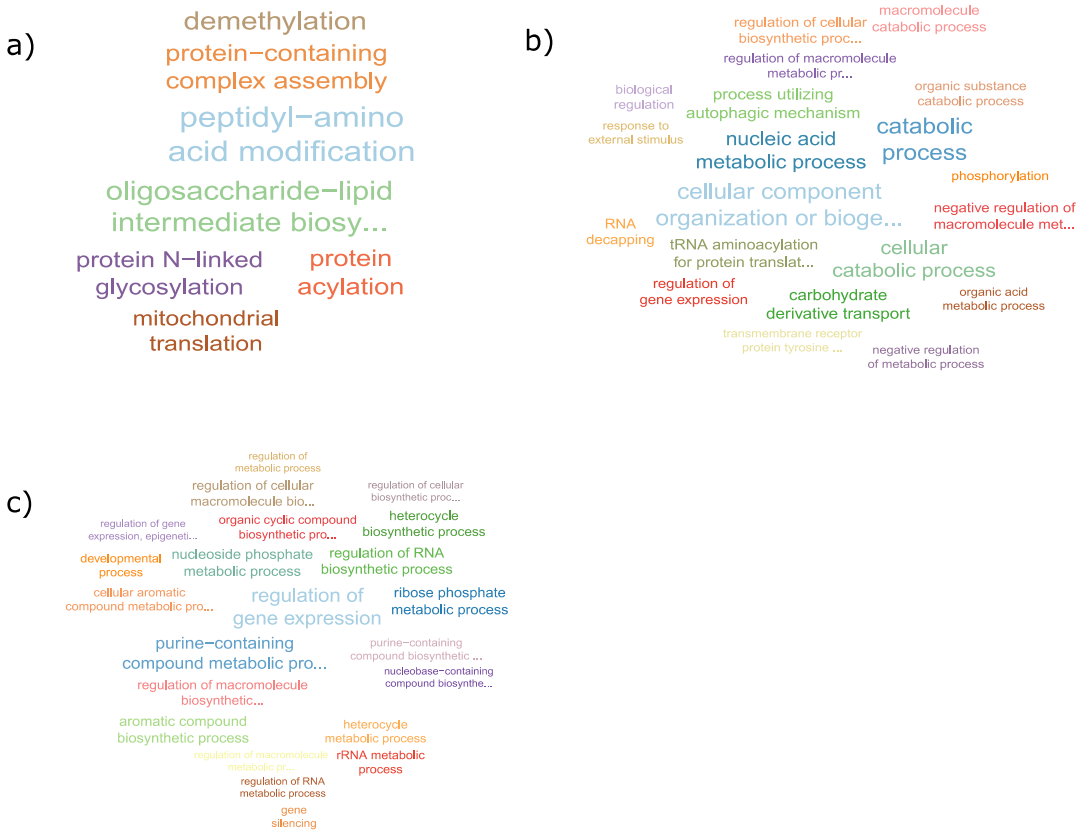

Figure 4: Word clouds of enriched GO functions based on genes with signature of selection in a) slavemakers, b) hosts, c) non-hosts.

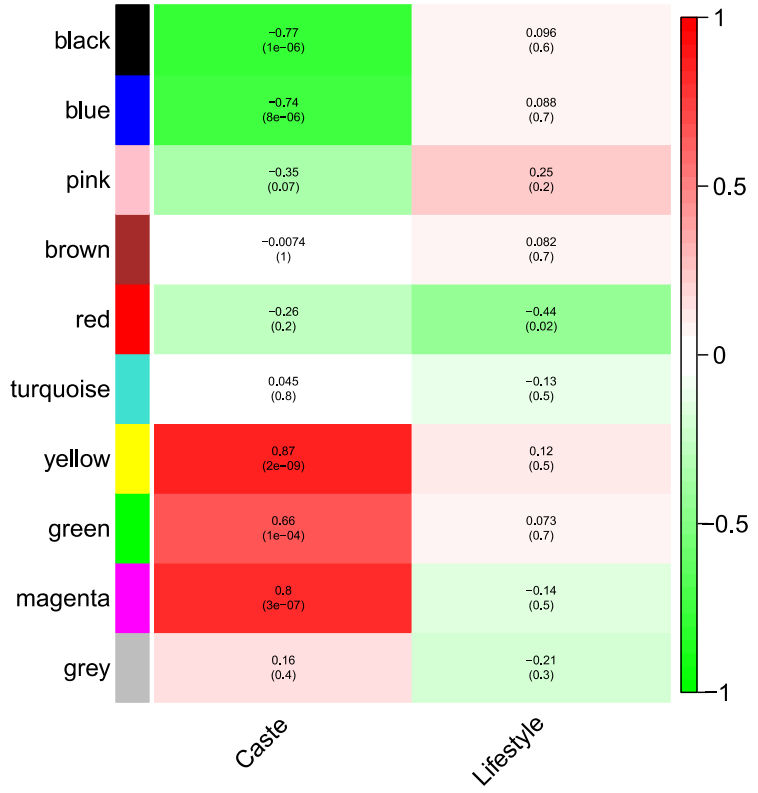

Figure 5: Module trait relationship of the 10 gene co-expression clusters with caste and lifestyle. The strength of the correlation is given in the upper numbers and significance levels in parentheses below. Modules are labelled by colours according to standard WGCNA output. Green indicates modules of genes overexpressed in workers / slave-makers, red indicates those with genes more expressed in queens / hosts.