

1 Changes in gene DNA methylation and expression networks accompany caste specialization
2 and age-related physiological changes in a social insect

3
4 Claire Morandin^{1,2,3*}, Volker P. Brendel^{4,5}, Liselotte Sundström^{1,6}, Heikki Helanterä^{1,7}, Alexander
5 S. Mikheyev^{8,9*}

6
7 **Affiliations:**

8 ¹ Organismal and Evolutionary Biology Research Programme, Faculty of Biological and
9 Environmental Sciences, University of Helsinki, Finland

10 ² Department of Biosciences, Åbo Akademi, Åbo, Finland

11 ³ Department of Ecology and Evolution, Biophore, University of Lausanne, 1015 Lausanne,
12 Switzerland

13 ⁴ Department of Biology, Indiana University, Bloomington, IN 47405, USA

14 ⁵ Department of Computer Science, Indiana University, Bloomington, IN 47405, USA

15 ⁶ Tvärminne Zoological Station, University of Helsinki, J.A. Palménin tie 260, FI-10900 Hanko,
16 Finland

17 ⁷ Ecology and Genetics research unit, Faculty of Science, University of Oulu, Finland

18 ⁸ Okinawa Institute of Science and Technology 1919-1 Tancha Onna-son, Kunigami-gun
19 Okinawa 904-0412, Japan

20 ⁹ Research School of Biology, Australian National University, Canberra, ACT 0200, Australia

21

22 *** Corresponding authors:**

23 **Claire Morandin**

24 claire.morandin@helsinki.fi

25 Organismal and Evolutionary Biology Research Programme, Faculty of Biological and

26 Environmental Sciences, University of Helsinki, Finland Tel: +358 2 941 57703

27 **Alexander S. Mikheyev**

28 alexander.mikheyev@anu.edu.au

29 Research School of Biology, Australian National University, Canberra, ACT 0200, Australia

30 Tel: +61 2 6125 2467

31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53

Abstract

Social insects provide systems for studying epigenetic regulation of phenotypes, particularly with respect to differentiation of reproductive and worker castes, which typically arise from a common genetic background. The role of gene expression in caste specialization has been extensively studied, but the role of DNA methylation remains controversial. Here, we perform well-replicated, integrated analyses of DNA methylation and gene expression in brains of an ant (*Formica exsecta*) with distinct female castes using traditional approaches (tests of differential methylation) combined with a novel approach (analysis of co-expression and co-methylation networks). We found differences in expression and methylation profiles between workers and queens at different life stages, as well as some overlap between DNA methylation and expression at the functional level. Large portions of the transcriptome and methylome are organized into ‘modules’ of genes, some significantly associated with phenotypic traits of castes and developmental stages. Several gene co-expression modules are preserved in co-methylation networks, consistent with possible regulation of caste-specific gene expression by DNA methylation. Surprisingly, brain co-expression modules were highly preserved when compared with a previous study that examined whole-body co-expression patterns in 16 ant species (Morandin et al. 2016), suggesting that these modules are evolutionarily conserved and for specific functions in various tissues. Altogether, these results suggest that DNA methylation participates in regulation of caste specialization and age-related physiological changes in social insects.

Keywords

55 Co-methylation network, co-expression network, ageing, phenotypic plasticity, caste

56 **Introduction**

57 DNA methylation is the most studied epigenetic mechanism and has been linked to variation in
58 gene regulation in mammals (Maunakea et al., 2010; Shukla et al., 2011), plants (Ecker & Davis,
59 1986; Zemach, Mcdaniel, Silva, & Zilberman, 2010; Zilberman, Coleman-Derr, Ballinger, &
60 Henikoff, 2008), and insects (Bewick, Vogel, Moore, & Schmitz, 2017; Kucharski, Maleszka,
61 Foret, & Maleszka, 2008; Li-Byarlay et al., 2013). DNA methylation regulates a wide range of
62 cellular processes, such as development or disease (Jirtle & Skinner, 2007; Lister et al., 2009;
63 Waterland & Jirtle, 2003; Wolffe & Matzke, 1999) and has been shown to affect processes such
64 as gene expression (Keshet, Yisraeli, & Cedar, 1985; Tate & Bird, 1993), genomic imprinting (E.
65 Li, Beard, & Jaenisch, 1993; Razin & Cedar, 1991), and transcriptional regulation (Eden &
66 Cedar, 1994; Schübeler, 2015).

67
68 Biological characteristics, functions, localization, and even presence of DNA methylation vary
69 greatly among taxonomic lineages (Colot & Rossignol, 1999). For example, in mammals, DNA
70 methylation primarily occurs at CpG dinucleotides (Cheng & Blumenthal, 2008), with 60% to
71 90% of all CpG sites methylated (Bird, 1986; Lister et al., 2009). DNA methylation is
72 accomplished by DNA methyltransferase 3 (DNMT3) and persists due to the activity of the
73 maintenance methyltransferase, DNMT1 (Cheng & Blumenthal, 2008; Goll & Bestor, 2005;
74 Kim, Samaranyake, & Pradhan, 2009). In mammals, DNA methylation typically occurs in gene
75 promoter regions, where it represses gene transcription (Bird & Wolffe, 1999; Suzuki & Bird,
76 2008; Weber et al., 2007). In insect species with identified DNA methylation machinery, DNA
77 methylation is predominantly found in coding regions and located in gene bodies of actively
78 expressed genes (Bonasio et al., 2012; Feng et al., 2010; Lyko et al., 2010; Suzuki & Bird, 2008;

79 Zemach et al., 2010). Gene body DNA methylation in insects correlates with alternative splicing
80 and may modulate gene activities (Bonasio et al., 2012; Flores et al., 2012; Foret et al., 2012;
81 Libbrecht, Oxley, Keller, Jan, & Kronauer, 2016; Lyko et al., 2010), or even affect gene function
82 through nucleosome stability (Hunt, Glastad, Yi, & Goodisman, 2013a). This suggests that gene
83 body DNA methylation may be involved in a wide range of biological processes, and that it may
84 be involved in evolution of novel traits, through mechanisms such as genomic imprinting
85 (Amarasinghe, Clayton, & Mallon, 2014). Furthermore, invertebrate genomes often display
86 lower levels of DNA methylation than those of mammals or plants, ranging from 0% (the order
87 Diptera) to 14% of all CpG sites (Asian cockroaches, *Blattella asahinai*) (Bewick et al., 2016).
88 In the honey bee (*Apis mellifera*), a model species for social insect genomic analyses, less than
89 1% of CpG dinucleotides are methylated (Lyko et al., 2010). Furthermore, in some insect species
90 (e.g. *Aedes aegypti*) no evidence of DNA methylation has been found. In these species, the
91 responsible machinery (cytosine-5 DNA methyltransferases) is absent from their genomes
92 (Falckenhayn et al., 2016; Standage, Berens, Glastad, & Severin, 2016). When present, DNA
93 methylation in insects contributes to diverse processes, such as nutritional control of
94 reproductive status (Kucharski et al., 2008), development (Lyko et al., 2010; Shi, Yan, & Huang,
95 2013; Yang, Guo, Zhao, Sun, & Hong, 2017), embryogenesis (Kay, Skowronski, & Hunt, 2017),
96 alternative splicing (Bonasio et al., 2012; Flores et al., 2012; Foret et al., 2012; Li-Byarlay et al.,
97 2013; Libbrecht et al., 2016), host-parasite evolution (Vilcinskis, 2016), memory processing
98 (Biergans, Jones, Treiber, Galizia, & Szyszka, 2012; Lockett, Helliwell, & Maleszka, 2010), age-
99 related changes in worker behavior (Herb et al., 2012), modulation of context-dependent gene
100 expression (Wedd, Kucharski, & Maleszka, 2016), maternal care (Arsenault, Hunt, & Rehan,
101 2018), and defense against territorial intrusion (Herb, Shook, Fields, & Robinson, 2018). Perhaps

102 most dramatically, in social insects, DNA methylation has been proposed to control the
103 developmental path taken by a totipotent egg to either a reproductive queen or a non-
104 reproductive worker (Herb et al., 2012; Kucharski et al., 2008; Yan, Bonasio, Simola, & Berger,
105 2015). Despite a common genetic background, queen and worker castes acquire extensive
106 behavioral, physiological and morphological differences, such as vast differences in lifespan,
107 which can be on the order of decades (Page & Peng, 2001). Because epigenetic regulation has
108 been proposed as a key mechanism in gene and environment interactions (Liu, Li, & Tollefsbol,
109 2008), social insects provide an ideal model to investigate the function of DNA methylation on
110 the development of alternative phenotypes (Lyko & Maleszka, 2011).

111

112 Since its initial discovery in honey bees (Wang et al., 2007), the role of DNA methylation in
113 regulating caste specification has been controversial. On one hand, many studies have reported
114 DNA methylation differences between female castes [(honey bees: Elango, Hunt, Goodisman, &
115 Yi, 2009; Foret et al., 2012; Kucharski et al., 2008; Lyko et al., 2010) (ants: Bonasio et al., 2012)
116 (bumblebees: Amarasinghe et al., 2014) (termites: Glastad, Hunt, & Goodisman, 2012; Glastad,
117 Gokhale, Liebig, & Goodisman, 2016)) and (non-social Jewel wasps, *Nasonia vitripennis*: Beeler
118 et al., 2014), while others found no effect of DNA methylation on caste regulation (Libbrecht et
119 al., 2016; Patalano et al., 2015). Several factors may explain these discrepancies. For instance,
120 studies that failed to detect significant DNA methylation differences compared reproductive and
121 non-reproductive individuals with similar morphology (Libbrecht et al., 2016; Patalano et al.,
122 2015). Also, most previous studies did not employ appropriate replication, and those that did,
123 failed to find significant differences between castes (Herb et al., 2012; Libbrecht et al., 2016;
124 Patalano et al., 2015). Thus, it is still unclear whether the lack of biological replicates or the lack
125 of distinct morphological castes explains the inconsistency among studies. Furthermore, only

126 two previous studies have used whole-genome sequencing to investigate DNA methylation
127 pattern differences between adult queen and worker brains in honey bees. Here again
128 discrepancies arise. Lyko et al. (2010), with no technical or biological replicates, found around
129 600 genes differentially methylated between castes, while Herb et al. (2012), with five replicates,
130 found no significant differences in DNA methylation between irreversible workers and queens.
131
132 Thus, to explore the role of DNA methylation on social insect caste and on individual traits
133 important in social organization, a study employing a suitable number of biological replicates
134 and a model system with clear caste differences was needed. To address this matter, we used the
135 ant, *Formica exsecta*, to study changes in brain DNA methylation and gene expression
136 associated with the two female castes. *F. exsecta* has morphologically differentiated castes, with
137 queens living as long as 20 years, while workers have lifespans slightly over 1 year, including a
138 winter hibernation (Pamilo, 1991). Thus, *F. exsecta* provides an extreme contrast in caste
139 physiology and lifespan, especially during the adult stage. In a previous study of the *F. exsecta*
140 transcriptome (Morandin et al., 2015), we found differential expression of DNA
141 methyltransferase 3 (DNMT3, up-regulated in adult workers compared to queens), an enzyme
142 responsible for establishing *de novo* DNA methylation patterns in mammalian genomes (Hata,
143 Okano, Lei, & Li, 2002; Kato et al., 2007; Okano, Bell, Haber, & Li, 1999; Okano, Xie, & Li,
144 1998), which also affects caste development in honey bees (Kucharski et al., 2008). We
145 hypothesized that DNA methylation states may differ between ant castes, either as a result of
146 differential DNA methylation during larval development, or because of re-programming as
147 adults. To test this hypothesis, we sampled queens and workers at two adult developmental
148 stages, soon after emergence from the cocoon and well after establishment in specific roles

149 (foraging and social behavior *vs.* reproduction). We then tested predictions that (a) both DNA
150 methylation and gene expression should differ between these stages; (b) if differences exist,
151 expression and DNA methylation signals would show some level of correspondence; and (c)
152 adults within a caste show persistent differences throughout life, suggesting an action of DNA
153 methylation beyond larval development. While our experimental design cannot prove causality
154 between DNA methylation and caste differentiation, it can show that caste-specific DNA
155 methylation patterns have the potential to underpin differences in caste and adult development,
156 hopefully spurring further functional investigation.

157

158 Here, for the first time in social insects, we examined the relationships between networks of co-
159 expressed and co-methylated genes. Co-methylation networks were first used to describe
160 relationships among methylation profiles (Busch et al., 2016; Eijk et al., 2012; Horvath et al.,
161 2012, 2016; Rickabaugh et al., 2015). Co-methylation analysis relies on the fact that adjacent
162 CpG sites can be co-methylated due to locally coordinated activities of methyltransferases or
163 demethylases. Weighted network methods (such as WGCNA) can be used on any high-
164 throughput, continuous, or semi-continuous datasets and preserves the continuous nature of co-
165 methylation information (Langfelder & Horvath, 2008; B. Zhang & Horvath, 2005) by grouping
166 highly correlated DNA methylation profiles into modules of genes. The overall DNA
167 methylation level of genes clustered in a module can be represented by the module eigengene
168 (Langfelder & Horvath, 2007; Langfelder, Mischel, & Horvath, 2013), which can later be
169 correlated with several phenotypic traits. In addition, robust preservation statistics are also
170 implemented in WGCNA and can be used to detect connections between modules of co-
171 expressed and co-methylated genes (Langfelder, Luo, Oldham, & Horvath, 2011).

172

173 Using DNA methylomes and expression profiling of individual queen and worker brains, we find
174 a number of differentially methylated genes and CpG sites associated with either caste or
175 developmental stages (newly emerged *vs.* old). In parallel, we find that the transcriptome and
176 methylome can be partitioned into conserved modules of co-expressed and co-methylated genes,
177 which are associated with caste and age-related physiological changes. Furthermore, some
178 methylation modules are preserved in the gene expression data, consistent with a possible
179 regulatory role of DNA methylation.

180

181 **Materials and Methods**

182 **Sample collection and brain extractions**

183 All samples of *F. exsecta* were collected from colonies around the Tvärminne Zoological Station
184 in the Hanko Peninsula, southwestern Finland, in the spring of 2013. Old adult queens and old
185 adult workers were collected in April, when ants come to the colony surface for warmth,
186 providing the only opportunity to easily collect egg-laying queens in the wild. The age of
187 overwintered queens could not be controlled; however, they were all found in large mature
188 colonies and were physogastric at the time of sampling (i.e., with greatly enlarged gasters due to
189 egg production). At this time of year, all workers have overwintered once and are reaching the
190 ends of their lives. Emerging queens were collected in June and emerging workers in July, right
191 after they emerged from their cocoons. Samples were collected randomly from 19 colonies in
192 close proximity, without bias toward specific ages or castes. After collection, samples were
193 frozen immediately at -80 °C. Brains were dissected on ice and stored in 180 µL buffer ATL and
194 20 µL proteinase K overnight (DNA, QIAamp DNA Micro Kit, Qiagen) or 350 µL buffer RLT
195 (RNA, RNeasy Micro Kit, Qiagen) for further extractions.

196

197 ***Formica exsecta* genome data**

198 *Formica exsecta* genome sequences and annotation (NCBI BioProject ID PRJNA393850 and
199 BioSample: SAMN07344805) were obtained pre-publication from the authors (Dhaygude, Nair,
200 Johansson, Wurm, & Sundström, 2018). In brief, the genome assembly consists of 14,617
201 contigs and scaffolds comprising a total of 278 Mb with an overall GC content of 36%.
202 Annotation of the genome reported 13,637 protein-coding genes, labeled FEX0000001 to
203 FEX0013637. Functional annotation was not provided, and classification of gene models for this
204 study was accomplished using BLAST (blast.ncbi.nlm.nih.gov/Blast.cgi) (Altschul, Gish, Miller,
205 Myers, & Lipman, 1990) and Blast2Go (www.blast2go.com/) (Conesa et al., 2005) searches, as
206 described below.

207

208 **DNA methylation analysis**

209 **DNA extraction, reduced representation bisulfite sequencing (RRBS) library preparation,**

210 **and sequencing.** RRBS was performed on 24 libraries from single ant brains (6 emerging

211 queens, 5 old queens, 6 emerging workers, 7 old workers), representing biological replicates of

212 each caste and developmental stage. Total DNA from each brain was extracted using a DNA

213 micro kit (QIAGEN) and diluted in 20 μ L of buffer AE (QIAGEN). Concentration and quality of

214 extracted DNA was examined with an Agilent 2100 bioanalyzer (Agilent Technologies).

215 *Restriction enzyme digestion.* DNA was digested with two enzymes (MspI and TaqI) prior to

216 bisulfite conversion (Gu et al., 2011). The following procedure was carried out for each enzyme

217 separately before combining samples at a later stage. First, 50 ng of DNA were mixed with 2 μ L

218 of 10 x T4 DNA ligase buffer, 1 μ L unmethylated lambda DNA (0.45 ng/ μ L), 0.5 μ L of the

219 enzyme (20 U/ μ L), and distilled H₂O up to 20 μ L. The mixture was then incubated at 37 °C
220 overnight.

221 *Adaptor ligation.* 1.25 μ L of methylated adaptor (15 μ M), 0.5 μ L of 10 x T4 DNA ligase buffer,
222 1 μ L of T4 DNA ligase (2,000 U/ μ L), and 2.25 μ L of distilled H₂O were added to each mixture.
223 Sequences of the methylated adaptors can be found in Table S1. The mixture was incubated at 4
224 °C overnight.

225 *Size selection.* Complete details of the size selection step can be found in Tan & Mikheyev
226 (2016). In brief, in the first selection step, 100 μ L of 13% PEG-6000/NaCl/Tris and 10 μ L
227 prepared Dynabeads were added to the mixture and resuspended. The mixture was incubated for
228 5 minutes and placed on a magnetic stand for 5 minutes. The supernatant (150 μ L) was
229 transferred to a new tube, and the beads were discarded. In the second selection step to select
230 fragments between 200 and 400bp, 100 μ L of 13.5 % PEG-6000/NaCl/Tris and 10 μ L prepared
231 Dynabeads were added to the supernatant and mixed. The mixture was incubated for 5 minutes
232 followed by bead separation on a magnetic stand. This time, the supernatant was discarded, and
233 the beads were collected. The beads were washed twice with 70% ethanol (with 10 mM Tris, pH
234 6) and dried for 5 minutes. The tubes were then taken off the magnetic stand, and DNA was
235 eluted from the beads by resuspending them in 15 μ L EB.

236 *Bisulfite conversion.* Unmethylated cytosines were converted to uracils using the Qiagen EpiTect
237 Bisulfite Kit. The kit was used twice on each sample following the manufacturer's instructions.

238 *Library amplification.* Bisulfite libraries were synthesized with a limited number of PCR cycles
239 (20). The 50 μ L PCR reaction consisted of 5 μ L of the bisulfite-converted genomic DNA from
240 the previous step combined with 5 μ L of 10 \times Advantage 2 PCR buffer (Clontech), 1.25 μ L of 10
241 mM dNTP mix, 2.5 μ L of 5 μ M Illumina primer, 1 μ L of 50x Advantage 2 DNA Polymerase

242 (Clontech) and 35.25 μ L of distilled H₂O. 50- μ L PCR reactions were set up for each bisulfite-
243 converted DNA sample. PCR reactions were carried out under the following conditions: initial
244 denaturation at 95 °C for 1 minute, with 20 cycles of denaturation at 98 °C for 10 seconds, 65 °C
245 for 1 min, followed by final extension at 72 °C for 5 minutes. PCR products were purified by
246 solid phase reversible immobilization using Dynabeads MyOne Carboxylic Acid (Invitrogen).
247 We followed the above protocol using 14.5% PEG for purification. Prior to pooling, libraries
248 were analyzed with a Bioanalyzer High-Sensitivity DNA Kit (Agilent Technologies). The
249 quantity of the library was estimated using a Quant-iT PicoGreen dsDNA Assay Kit and libraries
250 were pooled. Quantitative PCR (KAPA Biosystems) was used to estimate library concentrations.
251 Pooled libraries were sequenced single-end for 50 cycles (1 x 50bp) on an Illumina Hiseq 2000
252 system at the Okinawa Institute of Science and Technology. Adaptors were removed from the
253 raw reads. Subsequently the reads were parsed through quality filtration (Trimmomatic (Bolger,
254 Lohse, & Usadel, 2014), options: MAXINFO:40:0.8 MINLEN:10). Read quality was inspected
255 with FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Raw sequences were
256 submitted to the DDBJ database (see Table S2 for the accession numbers), ranging from 198 Mb
257 to 802 Mb of total sequence data per sample. Median trimmed read length for the samples
258 ranged between 33 and 46 nucleotides. Mapping efficiencies (next section) ranged between
259 64.3% and 73.3%, with median 69.5% (Table S2). To estimate the number of genomic cytosines
260 accessible to mapping after the RRBS protocol of fragmentation and size selection, we compared
261 whole genome mapping efficiencies with mapping efficiencies obtained against in silico
262 fragmented genomes. Results showed that mapping against fragments in the 50-440-bp range
263 gave almost the same mapping efficiencies as for mapping against the whole genome. Based on

264 this result, we estimate that about 11 million of the 24.5 million genomic CpGs were accessible
265 with our RRBS protocol.
266
267 **Computational processing of BS-seq data.** Individual read samples were mapped to the
268 *Formica exsecta* genome and DNA methylation calls were tallied using using Bismark (Krueger
269 & Andrews, 2011) with the BWASP workflow (<https://github.com/brendelgroup/BWASP>),
270 modified for RRBS data. In brief, BWASP is a workflow-enabling wrapper for Bismark, a BS-
271 seq analysis tools that executes read and mapping quality control and produces sets of highly-
272 supported DNA methylation (hsm) sites. hsm sites are a subset of sufficiently covered sites (scd),
273 i.e. genomic cytosines covered by enough reads to assess statistically significant DNA
274 methylation at that site. Requisite calculations are done in the BWASP Creport2Cxreport.py and
275 Cxreport2hsm.py scripts. Based on an assumed bisulfite conversion accuracy of 99.5%, a
276 binomial test is performed to determine whether the observed DNA methylation frequency at a
277 given site can be rejected as a chance event (1% significance level, Bonferroni adjusted). scd
278 sites are sites with sufficient coverage (here: 4 reads; Table S3) to allow detection of significant
279 DNA methylation, and hsm positions are where significant DNA methylation occurs (e.g., 4/4
280 DNA methylation calls). Overall levels of DNA methylation in CpG, CHG, and CHH sequence
281 contexts were estimated from mapped reads with BWASPR, as well as mean CpG methylation
282 levels of introns, exons, 5' UTRs, 3' UTRs and intergenic regions. We also calculated and
283 reported per-gene CpG methylation levels for queens, workers, emerging, and old samples.
284 Samples were compared only on the basis of sites that were covered in both samples when
285 replicates were averaged. Analysis of these sets of sites was done with a set of R functions that
286 are available in the BWASPR package (<https://github.com/brendelgroup/BWASPR>). A typical
287 BWASPR workflow reads the specified mcalls files (produced by Bismark) and generates

288 various output tables and plots, including differential DNA methylation analysis, as described
289 briefly below. Entire workflows are reproducible on any Linux system, following installation of
290 the packages. For convenience, we made all workflow documentation and scripts available on
291 the bgRAMOSE virtual machine (VM) image on the Jetstream scientific cloud computing
292 platform (<https://jetstream-cloud.org>). Users can deploy an example of this VM image and can
293 follow instructions from `/usr/local/share/bgRAMOSE/MBSHM2018/0README` to recap all of
294 our DNA methylation analysis workflows with a few keystrokes. Bismark's sam file output
295 (mcall) was used as input to methylKit (Akalin et al., 2012) and data were imported using the
296 function `read.bismark`. Differentially methylated CpG sites were determined with BWASPR
297 using logistic regression implemented in MethylKit (Akalin et al., 2012) from calls using the
298 functions `methylKit:calculateDiffMeth()` and `getMethylDiff`. For site-level analysis, we
299 discarded CpG sites covered by fewer than 10 reads, and we considered sites differentially
300 methylated if they showed 25% methylation differences and a qvalue of less than 0.01. pvalues
301 were adjusted to qvalues to account for multiple testing using the SLIM method. At the gene
302 level, differentially methylated gene lists for all four comparisons (OQ vs. OW; EQ vs. EW; OQ
303 vs. EQ; OW vs. EW) were compared among samples with the Wilcoxon paired ranked sign test,
304 applied only to genes of at most 20 kb and with at least two differentially methylated sites
305 (restrictions applied to focus on the genes with highest concentration of sites).

306

307 **Weighted correlation network analysis of DNA methylation.** We employed weighted co-
308 methylation networks analysis using the R package WGCNA (version 1.61.86, (Langfelder &
309 Horvath, 2008)) to find weighted signed co-methylated sets of genes (modules) associated with
310 caste and/or developmental stages, an approach analogous to that employed for gene expression
311 data. The goal of our network analysis was to 1) identify sets of co-methylated genes (modules),

312 2) calculate module eigengenes (i.e., representative values for each module), and 3) correlate
313 module eigengenes with phenotypes of interest (caste and stage). WGCNA identifies modules of
314 co-methylated genes starting at the level of DNA methylation and correlates these modules to
315 phenotypic traits. The network is created purely by gene DNA methylation levels and does not
316 require genes to be classified into binary categories (*i.e.*, whether a gene is methylated or not), as
317 is typical for gene-level differential DNA methylation tests. Thus, it overcomes the problem of
318 multiple comparisons. The input dataset (Table S4) consisted of results from DNA methylation
319 calls of 12,112 genes measured as the average percentage of CpG methylation per site per gene,
320 restricted to sites with highly-supported methylation data (high coverage, at least 10 reads).
321 These proportional levels control for the number of restriction enzyme sites present on each
322 gene, gene length, and quality of mapping to avoid any biases. WGCNA can be used on any
323 high-throughput continuous or semi-continuous data, and can calculate correlations from
324 proportional DNA methylation data without requiring normalization for gene length. Our dataset
325 was first filtered to remove genes with too many missing values, following WGCNA cutoff
326 threshold recommendations using the function `goodSamplesGenes` (Langfelder & Horvath,
327 2008). After considering a range of soft thresholding power (10 to 30), a power of 20 was chosen
328 based on the criterion of approximate scale-free topology and R^2 . After calculating topological
329 overlap values for all pairs of genes, a hierarchical clustering algorithm identifies modules of
330 highly interconnected genes. To define modules of co-methylated genes, we used average
331 linkage hierarchical clustering with the topological overlap-based dissimilarity measure.
332 Subsequently, modules of highly co-methylated genes were merged using a cut-off value of 0.45.
333 The minimum module size was set to 30 (Langfelder & Horvath, 2008). As detailed in Morandin
334 et al. (2016), we next calculated average signed, normalized gene DNA methylation values

335 (called an ‘eigengene’) to determine the relationship between modules and phenotypic traits
336 (e.g., caste and developmental stage). The eigengene is defined as the first principal component
337 of a module and represents the gene DNA methylation profile. For each module, the eigengene
338 can be used to define a measure of module membership, which indicates how close a DNA
339 methylation profile is to the module. A general linear model was then used to find the association
340 between external phenotypic traits (caste and developmental stage, and their interaction) and
341 modules’ eigengenes. The general linear model approach provides a convenient means of testing
342 the correlation of multiple traits with module eigengenes using a single model-relating eigengene
343 of expression or DNA methylation modules to caste phenotype and stage. We used the glm
344 function in R with 1000 bootstrap pseudoreplicates, with caste and stage as the explanatory
345 variables and their interactions. p-values were FDR-corrected to account for multiple testing.

346

347 **Gene expression analysis**

348 **RNA extraction, cDNA synthesis, and library preparation.** Four independent replicates for
349 each caste (queen, worker) and developmental stage (emerging, old), using single brains, were
350 used in this study. Total RNA from each brain was extracted using an RNeasy® micro kit
351 (QIAGEN) and diluted in 14 µL of RNase-free water. Concentrations and qualities of extracted
352 RNA were examined with an Agilent 2100 bioanalyzer (Agilent Technologies). Total input RNA
353 was standardized to 100 ng prior to cDNA synthesis. cDNA synthesis and library preparation
354 were done following an in-house protocol (Aird et al., 2013). Libraries were analyzed with a
355 Bioanalyzer High-Sensitivity DNA Kit (Agilent Technologies). Library quantities were
356 estimated with a Quant-iT PicoGreen dsDNA Assay Kit and equimolar concentrations of
357 libraries were pooled. Quantitative PCR (KAPA Biosystems) was used to estimate the

358 concentration of the libraries. Pooled libraries were sequenced paired-end with an Illumina
359 NextSeq High Output 2 x 150 bp (400M PE reads) at the FuGU lab in Helsinki (Finland).
360
361 **Read mapping and differential expression analysis.** Raw read quality was assessed with
362 FastQC tools (<http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc>), and adaptor sequences
363 were removed using cutadapt (Martin, 2011). Raw reads were subsequently parsed through
364 quality filtration (Trimmomatic (Bolger et al., 2014), options: MAXINFO:40:0.8 MINLEN:30),
365 and aligned to the reference *F. exsecta* genome (BioProject ID PRJNA393850, Dhaygude et al.,
366 2018) using Tophat2 (Kim et al., 2013) and Cufflinks (Trapnell et al., 2012). The genome
367 alignment output file was then used to reconstruct known transcripts using Cufflinks (Trapnell et
368 al., 2012). The transcriptome alignment output file of cuffmerge was subsequently used to
369 quantify expression levels of genes and transcripts using RSEM (Li & Dewey, 2011). The
370 resulting expected counts were used in the differential gene expression analysis with the R
371 Bioconductor package, EdgeR (Robinson, McCarthy, & Smyth, 2010). Reads generated from the
372 24 samples were used as replicates, and comparisons were made across castes and developmental
373 stages (OQ vs. OW; EQ vs. EW; EQ vs. OQ; EW vs. OW). For all comparisons, we first filtered
374 out transcripts with very low read counts by removing loci lower than 1 per kilobase of exons per
375 million fragments mapped in at least half of the sequenced libraries, as recommended by EdgeR.
376 TMM normalization was applied to account for compositional differences between libraries, and
377 expression differences were considered significant at a false discovery rate of $FDR < 0.05$.
378
379 **Weighted correlation network analysis of expression.** Weighted gene co-expression network
380 analysis was conducted using the R package WGCNA, as for the co-methylation network
381 detailed above. The input dataset consisted of a matrix with 16 libraries from either queens or
382 workers from both developmental stages, and 13,041 gene expression levels (Table S5). Log-

383 transformed FPKM values were used as input to avoid gene length biases, and as recommended
384 by Langfelder & Horvath (2008), the same procedure as in the co-methylation network analysis
385 was used to construct the co-expression network, with two exceptions. After considering a range
386 of soft thresholding power (10 to 30), a power of 20 was chosen, based on the criterion of
387 approximate scale-free topology and R^2 , and modules of highly co-expressed genes were merged
388 using a cut-off value of 0.2 (Langfelder & Horvath, 2008).

389

390 **Functional analysis and overlap between differentially methylated and differentially** 391 **expressed genes**

392 The software Blast2GO (www.blast2go.com) was used to infer functional annotation of the *F.*
393 *exsecta* gene set using structural similarity (BLASTx with an e-value cut-off $\leq 10^{-3}$). The Gostat
394 package for R (Beissbarth & Speed, 2004) was used to conduct GO term enrichment analysis on
395 differentially expressed and differentially methylated gene sets, using all genes having GO terms
396 as the universe. A similar procedure was used to conduct GO-term enrichment analysis on co-
397 expressed and co-methylated gene sets retrieved from WGCNA. Overlaps between differentially
398 expressed and differentially methylated gene lists were visualized using a Venn diagram
399 (<http://bioinformatics.psb.ugent.be/webtools/Venn/>) and statistical analysis of the significance of
400 overlaps between the two gene sets was calculated using the GeneOverlap BioConductor
401 package (Shen & Sinai, 2016).

402

403 **Association between DNA methylation and expression at the network level**

404 To assess preservation of modules between the expression and DNA methylation networks, we
405 used the network module preservation statistics Zsummary methods implemented in the R
406 function modulePreservation in the WGCNA R package (Langfelder & Horvath, 2008;

407 Langfelder et al., 2011). Network module preservation statistics assess whether the density
408 (strength of interactions among genes in a module) and connectivity patterns of modules (for
409 example hub/central genes) are preserved between two independent datasets. This method does
410 not require that modules and genes be identified in the target network (only in the reference data
411 set); therefore, it is independent of ambiguities associated with module identification (Langfelder
412 et al., 2011). The Zsummary statistic result summarizes the evidence that network connections of
413 the modules are significantly more preserved than connections of random sets of genes of equal
414 size (Langfelder et al., 2011). We used the ‘modulePreservation’ function in the WGCNA R
415 package (Langfelder & Horvath, 2008) that enables rigorous testing of module preservation with
416 100 permutations. This reproducibility method was used to estimate the relationship between
417 modules obtained from our co-expression and co-methylation network. In addition, to understand
418 the biological relevance and reproducibility of our analysis, we conducted the same module
419 preservation statistics (Langfelder et al., 2011) between our co-expression network modules and
420 modules retrieved from a recent study of caste specialization across 16 ant species (Morandin et
421 al., 2016).

422 Detailed scripts can be found in <https://github.com/oist/formica-methylation> under a MIT
423 license.

424

425 **Results**

426 **DNA methylation profiles**

427 *Overall patterns:* DNA methylation frequency was assessed in two ways. The overall level of
428 DNA methylation in the mapped reads (e.g. as reported by Bismark (Krueger & Andrews, 2011)
429 and referring to all DNA methylation calls made on all mapped reads) was found to be about 1%

430 in all samples. The proportions of highly supported DNA methylated (hsm) CpGs in the
431 respective sets of detectable (scd) sites were 2.93% for emerging queens, 1.63% for emerging
432 workers, 1.71% for old queens and 1.95% for old workers (Table S6 and S7; Figures S1, S2).
433 Hsm statistics/proportions allow us to establish whether DNA methylation is localized. These
434 low DNA methylation levels are similar to previous observations for honeybees (0.1%, Lyko et
435 al., 2010), the ants, *Camponotus floridanus* (0.3%, Bonasio et al., 2012), *Harpegnathos saltator*
436 (0.2%, Bonasio et al. 2012), and *Cerapachys biroi* (2.1%, Libbrecht et al. 2016). Consistent with
437 observations on other hymenopterans, cytosine DNA methylation was found almost exclusively
438 in a CpG context (Table S8). Highly supported methylated DNA CpGs sites occur in all
439 annotated genomic regions, but are observed at about 2.5-fold higher numbers in coding regions
440 than expected by random distribution over all scd sites (Tables S9 and S10). DNA methylation
441 was found mostly within genes (90.8% +/- 3.5%), and predominantly in exons (64.5% +/- 8%)
442 (Figure 1, Tables S9 and S10). There was a positive correlation between expression levels and
443 the mean % DNA methylation per site, implying that highly expressed genes are more likely to
444 be highly methylated ($cor = 0.28$, $p < 0.001$; Spearman's product-moment correlation, Figure 2).
445 Results of the Wilcoxon rank-sum test showed that overall DNA methylation levels were
446 significantly higher in queens compared to workers ($p < 0.001$), at both emerging and old stages.
447 Emerging ants had higher overall DNA methylation than old ants among both queens and
448 workers ($p < 0.001$ each; Wilcoxon signed rank test).
449 As a control for our experiments, lambda DNAs were included to rule out any issues with
450 conversion efficiency differences between samples. Conversion rates were verified by mapping
451 the BS-reads to the lambda genome. None of the samples show any conversion problems. For
452 instance, for the 6 emerging worker samples, the % of unmethylated C calls on lambda DNA by

453 Bismark were between 99.38% and 99.71%. Our sequencing depth was not sufficient to
454 meaningfully assess and compare the DNA methylation status of replicates individually. Rather,
455 the statistical analysis relies on comparisons made between the replicated aggregate groups with
456 safeguards against accidental reliance on a non-representative individual.

457 *Differentially methylated sites:* Across castes, and after correcting for multiple testing, we found
458 1,528 sites differentially methylated between old queens and old workers (869 up-methylated in
459 old workers and 659 in old queens), and 1,620 differentially methylated sites between emerging
460 queens and emerging workers (886 sites up-methylated in workers, and 734 in queens).

461 Similarly, between developmental stages, we found 1,344 sites differentially methylated between
462 emerging and old queens (692 sites up-methylated in emerging queens and 652 sites up-
463 methylated in old queens). Likewise there were 1,894 sites differentially methylated between
464 emerging and old workers (949 sites up-methylated in emerging workers and 945 sites in old
465 workers) (Table S11).

466
467 *Differentially methylated genes:* At the gene level and across castes, these differentially
468 methylated CpG sites resulted in 226 genes (1.7%) having at least two sites differentially
469 methylated between old queens and old workers (118 genes up-methylated in queens and 108 in
470 workers), with no significant differences in the number of differentially methylated genes ($p =$
471 0.55, Fisher's exact test). Between emerging queens and emerging workers, 264 genes (1.9%)
472 were differentially methylated with 164 genes up-methylated in emerging queens and 100 genes
473 in emerging workers. Emerging queens up-methylated significantly more genes than emerging
474 workers ($p < 0.001$, Fisher's exact test, Figure S3). 75 genes were differentially methylated
475 between castes in both old and emerging stages, which is more than expected by chance (Figure
476 S4, GeneOverlap R package, $p < 0.001$). Over half of those genes were caste-biased in opposite

477 directions across both developmental stages (41 genes, 55%). These results parallel expression
478 patterns, suggesting that caste-biased genes are more specific to developmental stage than to
479 caste.

480 Across stages, 198 genes (1.5%) had at least two sites differentially methylated between
481 emerging queens and old queens (105 sites up-methylated in emerging queens and 93 in old
482 queens). No significant differences were found in the number of genes up-methylated in
483 emerging and old queens ($p = 0.43$, Fisher's exact test). And 300 genes (2.3%) had at least two
484 sites differentially methylated in emerging and old workers (172 genes up-methylated in
485 emerging and 128 in old workers). A significant difference was found in the number of genes up-
486 methylated between emerging workers and old workers ($p = 0.01$, Fisher's exact test, Figure S3).

487 There were 66 genes in common among those differentially methylated between stages for
488 queens and workers; however, this is not more than expected by chance (Figure S4, GeneOverlap
489 R package, $p = 0.2$). However, half of those genes were over-methylated in different direction
490 (35 genes, 53%). Blast annotations of the differentially methylated gene lists for all comparisons
491 can be found in Table S12.

492

493 *Differential DNA methylation GO term annotation:* The complete list of enriched GO terms can
494 be found in Table S13, but here we summarize some of the most conspicuous findings. Between
495 old queens and old workers, the queen up-methylated gene list included GO terms such as
496 response to stress and DNA repair, whereas worker up-methylated genes were associated with
497 oxoacid metabolic process and ncRNA metabolic process. Between emerging queens and
498 emerging workers, the queen up-methylated gene list included GO terms similar to the old stage,
499 such as DNA repair and ncRNA metabolic process. Worker gene list included GO terms such as
500 cellular response to DNA damage stimulus and positive regulation of catabolic processes. Across

501 queens, old ants up-methylated genes that were related to GO terms such as ncRNA metabolic
502 processes and DNA repair and methylation, whereas visual perception and eye morphogenesis
503 were enriched in the emerging ant up-methylated gene set. Across workers, old ants up-
504 methylated genes that were related to GO terms such as oxoacid metabolic processes and ATP
505 metabolic processes. In contrast, RNA processing and DNA replication initiation were enriched
506 for the emerging ant up-methylated gene set. Sets of GO terms associated with caste and/or
507 developmental stage differences are consistent with hypothetical regulatory roles for differential
508 DNA methylation (similar to previous studies (Foret et al., 2012; Kucharski et al., 2008;
509 Libbrecht et al., 2016; Lyko et al., 2010).

510

511 **Expression profiles**

512 *Overall patterns:* We recovered 99 Gb of 100-bp paired-end reads from the 16 libraries.
513 Following quality filtering, we realigned the reads to the *F. exsecta* genome, and on average a
514 mapping rate of ~84% was obtained. A total of 10,874 genes were expressed, with over 1 count-
515 per-million in at least half of the samples (Robinson, McCarthy, & Smyth, 2010).

516

517 *Differential expression patterns:* Among old queens and old workers, a total of 1,185 genes
518 (8.7% of the total number of genes present in the genome) were differentially expressed, with
519 queens up-regulating significantly more genes than workers (675 vs. 510, $p < 0.001$, Fisher's
520 exact test). Between emerging queens and emerging workers, a total of 416 genes (3.1%) were
521 differentially expressed, with queens over-expressing more genes than workers (291 vs. 125, $p <$
522 0.001 , Fisher's exact test) (Figures S3, S5). We compared the list of genes differentially
523 expressed between castes across old and emerging stages and found 164 genes overlapping.
524 These genes are always differentially expressed between queens and workers regardless of the
525 developmental stage (Figure S4, GeneOverlap R package, $p < 0.001$). However, many genes (48

526 genes, 29.3%) were caste-biased in opposite directions in different developmental stages,
527 meaning that caste-biased genes tend to be specific to developmental stages. A similar pattern
528 was also found in a previous study looking at caste-biased genes over several development stages
529 in the same species (Morandin et al., 2015).

530 Across stages, within the queen caste, a total of 892 genes (6.5% of all genes in the genome)
531 were found differentially expressed between emerging and old ants, with no significant
532 differences in the number of genes over-expressed (433 vs. 459, $p = 0.4$, Fisher's exact test).
533 Within the worker caste, a total of 1,568 genes (11.5%) were differentially expressed between
534 emerging and old workers, with more genes up-expressed by the emerging workers (893 vs. 675,
535 $p < 0.001$, Fisher's exact test, Figure S3, S5). We compared the genes differentially expressed
536 between stages across queen and worker castes, and found 292 common genes. These genes are
537 always differentially expressed between stages regardless of caste (Figure S4, GeneOverlap R
538 package, $p < 0.001$). Many genes were consistently over-expressed by the same caste (269 genes,
539 92.1%) between developmental stages. Blast annotations of the differentially expressed gene lists
540 for all comparisons can be found in Table S12.

541
542 *Differential expression GO term annotation:* The complete list of enriched GO terms can be
543 found in Table S13. Between old queens and old workers, GO terms enriched for oxidation-
544 reduction process and hormone transport were associated with queens, and terms such as social
545 behavior and multi-organism behavior were enriched for workers. Between emerging queens and
546 emerging workers, GO terms such as oxidation-reduction process and response to hormone were
547 associated with queens, and cellular response to stimulus for workers. Between old and emerging
548 queens, old ants up-regulated genes for telomere organization and response to stress, while
549 emerging ants up-regulated genes associated with GO terms such as oxidation-reduction process

550 and regulation of TOR signaling. Comparing old and emerging workers, old ants enhanced
551 expression of gene for DNA recombination and sensory perception of smell, while emerging ants
552 over-expressed genes associated with regulation of hormone levels and response to stimulus.

553

554 **Overlap between differentially expressed and differentially methylated genes**

555 To examine the hypothesis that expression and DNA methylation signals would show
556 correspondence when looking at caste and developmental stage differences, we investigated
557 whether list of differentially expressed and differentially methylated genes (all comparisons)
558 might overlap at three different levels: genes and GO terms, in addition to network-based
559 analyses described below. When comparing the lists of genes, very few were both differentially
560 expressed and differentially methylated. We found 19 genes differentially expressed and
561 differentially methylated between old queens and old workers, and 6 genes between emerging
562 queens and emerging workers. We also found 8 genes that are both differentially expressed and
563 differentially methylated between emerging and old queens, and 37 genes between emerging and
564 old workers (Figure 3). The lists of genes differentially expressed and differentially methylated
565 genes across the four comparisons did not overlap significantly for any comparisons
566 (GeneOverlap R package, Old queens *vs.* workers (1185 *vs.* 226, $p = 0.6$); Emerging queens *vs.*
567 workers (416 *vs.* 264, $p = 0.82$); Emerging queens *vs.* old queens (892 *vs.* 198, $p = 0.95$);
568 Emerging workers *vs.* old (1568 *vs.* 300, $p = 0.35$); Figure 3). We also examined whether the
569 direction of up/down expression and DNA methylation (e.g. whether a gene that is more
570 expressed in one caste is also more methylated in the same caste). Surprisingly the direction of
571 overexpression/DNA methylation only matched in around half of the genes (Figure S6; OQ *vs.*
572 OW: 9 genes out of 19; EQ *vs.* OQ: 4 genes out of 8; EW *vs.* OW: 15 out of 37), except in the

573 case of emerging queens *vs.* emerging workers, where all genes (8), both differentially expressed
574 and differentially methylated, were upregulated by the emerging queens.

575
576 Next, we investigated whether we could find a correspondence between expression and DNA
577 methylation and gene function, looking at the overlap between lists of GO terms. Surprisingly,
578 given the small number of genes, we found a significant overlap between lists of GO terms
579 associated with genes differentially expressed and differentially methylated genes across the four
580 comparisons (GeneOverlap R package, Old queens *vs.* workers (158 *vs.* 162, $p < 0.001$);
581 Emerging queens *vs.* workers (51 *vs.* 243, $p < 0.001$); Emerging queens *vs.* old queens (68 *vs.*
582 137, $p < 0.001$); Emerging workers *vs.* old (136 *vs.* 264, $p < 0.001$); Figure 4 and Table S13).
583 Despite possible limitations with GO terms analyses, which rely on orthology with distantly
584 related references, they provide insights into biological processes possibly involved, beyond
585 what can be gleaned from gene lists alone.

586

587 **DNA methylation and expression modules correspond to castes and developmental** 588 **stages**

589 We separately constructed co-expression and co-methylation networks from the expression and
590 DNA methylation datasets using the Weighted Correlation Network Analysis approach
591 (WGCNA (Langfelder & Horvath, 2008)). In the methylation dataset, 8,200 genes were retained
592 for further analyses following the cleaning step. Due to low coverage, one old queen sample and
593 three old worker samples were removed from the co-methylation network input dataset. A total
594 of 6 emerging queens, 6 emerging workers, 4 old queens and 4 old workers were used for this
595 analysis. A total of 348 genes were not co-methylated and were excluded from further analysis.
596 A total of 20 co-methylated modules (labelled M1-M20) were identified, ranging in size from 56

597 (M1) to 3,230 (M18) genes, with 393 genes per module on average. In the expression dataset,
598 10,700 genes were retained after removing genes with too many missing values, and 6,570 genes
599 were subsequently assigned to one of the co-expression modules. A total of 4130 genes were not
600 co-expressed and were consequently removed from further analysis. We identified 14 co-
601 expression modules (labeled E1-E14) with sizes ranging from 46 (E11) to 2,454 (E9) genes, with
602 an average of 469 genes per module. For both datasets, we calculated the module eigengenes
603 which are defined as the first principal component of a module and are representative of gene
604 expression or gene DNA methylation profiles in a module (Langfelder & Horvath, 2008).

605 Afterward, we correlated these eigengenes with two phenotypic traits, *i.e.*, caste and stage, using
606 a glm approach. When the eigengene of a module is correlated with a trait of interest, it means
607 that most/all genes in the module exhibit a significant correlation/association with the trait, and
608 we can define which genes/modules are likely to underlie the phenotype via gene expression. For
609 the methylation dataset, one of the modules was significantly correlated with one of the caste
610 phenotypes (M2; Queen phenotype), and 11 of 20 modules were significantly correlated with
611 stage (Emerging phenotype). For the expression dataset, 7 out of 14 modules were correlated
612 with one of the two female castes (4 associated with the worker phenotype and 3 with the queen)
613 and 9 modules were correlated with stage (5 with the old phenotype and 4 with the emerging). In
614 addition, 3 methylation modules and 5 expression modules were significantly associated with the
615 interaction of caste and stage phenotypes (Figure 5, Table 1).

616 To gain insight into the biological relevance and functional significance of modules, we
617 performed GO term enrichment analysis on the genes in each module (Tables S14 and S15).
618 Here we summarize some of the main findings. In the methylation dataset, the module associated
619 with the queen phenotype (M2) was correlated with gene expression and RNA metabolic

620 processes. Modules associated with the emerging phenotype were correlated with gene
621 expression (M3), response to pheromone (M7), ATP biosynthetic process (M8), RNA metabolic
622 process (M11), mRNA transport (M12), oxidative phosphorylation (M14), nerve development
623 (M15), regulation of RNA biosynthetic process (M16), social behavior (M17), developmental
624 process (M18) and DNA repair (M19) (Table S15). In the expression dataset, queen-associated
625 modules were related to detection of chemical stimulus involved in sensory perception of smell
626 (E7), oxoacid metabolic process (E13) and oxidation-reduction process (E14). Worker-
627 associated modules were linked to social behavior (E1), TOR signaling (E2), response to stress
628 (E3) and sensory perception (E5). Modules associated with stage in the expression dataset, were
629 correlated with DNA recombination (E4), detection of chemical stimulus (E5), growth (E6),
630 muscle contraction (E7) and development of the central nervous system in old adults (E8), while
631 modules associated with the emerging phenotype were associated with biological functions such
632 as response to growth factor (E11), regulation of cell death (E12), cell cycle (E13) and response
633 to hormone (E14) (Table S15). As a precaution, it is worthy to note that GO terms associated
634 with eye pigmentation (M14) were also enriched and could potentially be a sign of
635 contamination.

636

637 **Preservation of co-expression sets of genes in DNA methylation data**

638 Next, we looked for evidence of preservation between the co-expression and co-methylation
639 networks using the WGCNA R package. Values of “*Zsummary* below 2” indicate no evidence of
640 preservation. Values between 2 and 5 indicate moderate evidence for preservation, while values
641 over 5 indicate strong evidence of preservation. Although conservation of modules between co-
642 expression and co-methylation datasets was weak, four expression modules were conserved in
643 the co-methylation data. Based on the preservation statistic *Zsummary*, we found that four co-

644 expression sets of genes (E1, E6, E9, and E10) exhibited moderate preservation within the co-
645 methylation network (Figure 6). The expression module E6 is correlated with the old phenotype,
646 module E9 with the association of both phenotypes, and module E1 with the worker phenotype
647 (Table S16). An online resource has been created to simplify visualization of module
648 organization and the association between expression and DNA methylation networks (the
649 website is available at <https://mikheev.github.io/formica-brain-expression-methylation/>).
650 The online tool allows users to visualize caste and stage association for each module (Figure 5).
651 Many of the modules belonging to the co-expression or co-methylation networks were correlated
652 with either caste (inner band), or stage (outer bands).

653

654 **Preservation of co-expression sets of genes with a previous study**

655 Similarly, we looked for preservation between our brain co-expression network and our multiple
656 species co-expression network published in an earlier study (Morandin et al., 2016). Based on
657 the preservation statistic Zsummary, we found that 3 out of 36 of our 2016 co-expression
658 modules exhibited moderate preservation with our brain co-expression network. Additionally, 9
659 out of 14 brain co-expression modules in return exhibited moderate to strong preservation with
660 the 2016 co-expression network (e.g. modules E1, E2, E3, E6, E8, E9, E10, E12 and E13)
661 (Figure S7, Table S17). This module preservation analysis confirmed that our modules are found
662 in an independent data set.

663

664 **Discussion**

665 The goal of this study was to examine whether socially important phenotypic traits correlate with
666 divergence in DNA methylation and expression. We found distinct transcriptional and

667 methylation differences between castes within a developmental stage and different
668 developmental stages within a caste. Furthermore, there was some evidence of overlap between
669 methylation and gene expression states, but only at the functional level, that of biological
670 processes gene ontology terms. More specifically, some modules in the co-expression data are
671 preserved in co-methylation modules (retain similar network structure (genes) and network
672 properties), though not *vice versa*. This is consistent with a role of DNA methylation as a
673 proximate mechanism regulating gene expression, which is already known to affect caste
674 specialization, as well as task specialization and ageing, in social insects (Kozeretska, Serga,
675 Koliada, & Vaiserman, 2017; Yan et al., 2015), though further investigations are necessary to
676 confirm this link.

677
678 Here we used individual brains to conduct the first reduced representation bisulfite sequencing
679 study on social insects. RRBS brings down the scale and cost of whole genome bisulfite
680 sequencing by only analyzing a representative portion of the genome. In vertebrates, this
681 approach has been shown to capture around 85% of the CpG islands and 60% of promoters while
682 requiring very little input material (Gu et al., 2011), and it permits more replicates per
683 experiment while providing an efficient way to generate overall quantification of DNA
684 methylation, and to apply powerful network-based analysis methods with individual-relevel
685 replication. Nonetheless, it is important to note the limitations of RRBS. Because restriction
686 enzymes digest DNA sequences at restriction sites randomly across the genome, many relevant
687 methylated fragments in each sample are missed. Thus, these results encompass only a fraction
688 of all methylated sites. As a result, global patterns are more easily captured compared to specific
689 mechanisms. Furthermore, the size selection step associated with RRBS results in stochastically

690 uneven coverage across samples, and large amounts of missing data. This must be carefully
691 accounted for in the analysis, and it also makes per-site comparisons, and comparisons across
692 more than two conditions problematic. In this analysis, after rigorously filtering sites based on
693 coverage and quality, we conducted analyses aggregated at the gene level. Despite these
694 limitations, RRBS produces data that are comparable across different experimental conditions,
695 and can provide biological insights, given the right statistical approach. In particular, co-
696 methylation analysis overcomes some of the limitations introduced by data sparseness, as DNA
697 methylation levels at nearby CpG-sites tend to be highly correlated. For instance, variation in
698 DNA methylation across treatments has been showed to occur more frequently in aggregated
699 CpGs (Jaffe, Feinberg, Irizarry, & Leek, 2012). Because co-methylation analysis leverages
700 information from the entire data set to construct the network, sporadic missing sites do not
701 significantly alter the overall data structure. Thus, this type of analysis is more powerful than
702 those focusing on detecting single-site differences, and allows more sophisticated types of
703 analyses, such as full-factorial designs with main effects and interaction terms (Table 1).

704

705 Previous work proposed that social insect castes differ in DNA methylation states, which are
706 established during caste differentiation that occurs during larval development (Bonasio et al.,
707 2012; Elango et al., 2009; Foret et al., 2012; Glastad et al., 2016; Kucharski et al., 2008; Lyko et
708 al., 2010), though that view has been challenged (Herb et al., 2012; Libbrecht et al., 2016;
709 Patalano et al., 2015). We tested this hypothesis, also expanding it to include changes in DNA
710 methylation states that take place in the course of adult development and ageing. To do this we
711 sampled queens and workers at different ages, and made comparisons both within castes across
712 developmental categories, as well as between castes within the same age category, without

713 maintaining that “old” and “emerging” are necessarily equivalent states for the two castes.
714 Indeed, emerging and old ants were sampled a few weeks apart, and emerging queens are also
715 unmated compared to old queens. Keeping in mind that age categories are difficult to standardize
716 between castes, we found significant changes in DNA methylation and gene expression between
717 these phenotypic endpoints, showing that DNA methylation differs between ant castes, as well as
718 across the lifetime of the adult. The latter finding is particularly interesting since it suggests that
719 caste differences due to DNA methylation may be dynamic, and not necessarily fully fixed
720 during the developmental program when the castes differentiate. The same pattern holds true
721 when comparing DNA methylation between castes over the two developmental stages,
722 suggesting that age-related changes in DNA methylation differ between queens and workers at
723 least at the adult stage. Developmental specificity of DNA methylation bias in general is
724 certainly a promising direction for future research. Furthermore, we found some significant
725 interaction terms in co-expression and co-methylation analysis, suggesting that gene expression
726 and DNA methylation levels do not necessarily change in the same direction as a function of
727 caste and adult developmental stage (Table 1). This suggests that ants have a dynamic DNA
728 methylation system that is active throughout their lives. Indeed, the overall positive relationship
729 between gene body DNA methylation and expression (Figure 2), which is a hallmark of other
730 insects, suggests that DNA methylation in ants largely functions in the same way as in other
731 species. From this perspective, finding differences between castes and developmental stages, as
732 we did in this study, is not surprising given the many ancient developmental and regulatory
733 mechanisms coopted into caste differentiation (Robinson, Grozinger, & Whitfield, 2005).
734

735 **Patterns of DNA methylation.** We found more overall DNA methylation in queens compared to
736 workers, at both developmental stages, in contrast to previous findings that found a lower level
737 of DNA methylation in queens during the larval stage (Shi et al., 2013), or even no differences in
738 the adult stage (Lyko et al., 2010) in the honey bee. However, similar results were observed for
739 the ant, *Pogonomyrmex barbatus*, in which virgin queens exhibited higher levels of DNA
740 methylation compared to workers (Smith et al., 2012). Unfortunately, studies are still too few to
741 draw any general patterns. Also, considering that *Formica exsecta* queens live ~20 years and the
742 workers just over a year (Pamilo, 1991), if DNA methylation accumulated over time/ageing, we
743 would expect to find large differences in overall DNA methylation between old queens and
744 emerging queens, and more subtle differences between emerging and old workers. However,
745 contrary to this hypothesis, we found the opposite pattern, with emerging ants showing more
746 DNA methylation overall than old ants. The important role of DNA demethylation in diverse
747 biological processes by regulating gene expression has been well documented in mammals
748 (Richardson, 2003), but its exact role with gene body DNA methylation remains unclear. Further
749 experimental studies of DNA demethylation in social insects (especially considering caste
750 longevity and behavior differences) are needed to further understand how DNA demethylation is
751 transduced into physiological changes over time in the two castes.

752
753 Previous work has demonstrated the usefulness of network-based approaches for detecting links
754 between expression and DNA methylation that could be missed by more commonly used
755 approaches focused on comparing lists of differentially expressed and differentially methylated
756 genes (Davies et al., 2012; Eijk et al., 2012; Horvath et al., 2012; J. Zhang & Huang, 2017). Our
757 results reveal that both the brain transcriptome and methylome can be organized into modules.
758 All co-methylation and co-expression modules are significantly enriched with gene ontology

759 categories (Table S14), thus providing additional evidence that these modules are biologically
760 meaningful. For instance, co-expressed modules associated with castes were enriched for gene
761 functions such as social behavior, TOR pathways and pheromone synthesis, while co-
762 methylation modules were enriched for core biological functions such as DNA repair.
763 Interestingly, co-expression modules identified in this study in brain tissue were also conserved
764 in whole-body transcriptional data from an earlier study (Figure S7), suggesting that similar gene
765 regulatory processes act at both tissue-specific and whole-body levels.

766

767 **Conclusion.** We propose that in addition to action on single genes or their isoforms, gene DNA
768 methylation may be thought of in a network context, with co-methylation modules associated
769 with specific phenotypes, *e.g.*, caste and stage (Table 1). We hope that future work will focus on
770 reconstructing ever more accurate co-methylation networks, which will require large numbers of
771 replicates across different phenotypic states to fully understand the role of DNA methylation and
772 how it interacts with gene co-expression to generate phenotypic novelty, as it has been done
773 recently for gene co-expression networks (Morandin et al., 2016). Investigations of diverse taxa
774 using similar methodologies would be particularly useful for identifying the extent to which
775 DNA methylation is associated with caste or other phenotypic traits among social insects, and
776 how it evolves.

777

778 **Acknowledgements**

779 This work was supported by the Otto A. Malmin lahjoitusrahasto grant (to CM), by the Academy
780 of Finland (grant numbers 140990,135970, and 273029 to HH, and 252411, 284666 to the Centre
781 of Excellence in Biological Interactions), by the University of Helsinki, by the Kone Foundation

782 (to HH), by the Okinawa Institute of Science and Technology Graduate University, and JSPS
783 KAKENHI nos. 24770034 and 25221206 (to ASM) as well as ARC grants DP170100772 and
784 FT160100178 (to ASM). We are grateful to Steve Aird for proofreading the manuscript.

785

786 **References**

787 Aird, S. D., Watanabe, Y., Villar-briones, A., Roy, M. C., Terada, K., & Mikheyev, A. S. (2013).
788 Quantitative high-throughput profiling of snake venom gland transcriptomes and proteomes
789 (Ovophis okinavensis and Protobothrops flavoviridis). *BMC Genomics*, *14*, 790.

790 Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F. E., Figueroa, M. E., Melnick, A., &
791 Mason, C. E. (2012). MethylKit: a comprehensive R package for the analysis of genome-
792 wide DNA methylation profiles. *Genome Biology*, *13*(10). doi:10.1186/gb-2012-13-10-R87

793 Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local
794 alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410.

795 Amarasinghe, H. E., Clayton, C. I., & Mallon, E. B. (2014). Methylation and worker
796 reproduction in the bumble-bee (*Bombus terrestris*). *Proceedings of the Royal Society B*,
797 *281*(1780), 3–8.

798 Arsenault, S. V., Hunt, B. G., & Rehan, S. M. (2018). The effect of maternal care on gene
799 expression and DNA methylation in a subsocial bee. *Nature Communications*, *9*(1), 3468.
800 doi:10.1038/s41467-018-05903-0

801 Beeler, S. M., Wong, G. T., Zheng, J. M., Bush, E. C., Remnant, E. J., Oldroyd, B. P., &
802 Drewell, R. A. (2014). Whole-genome DNA methylation profile of the jewel wasp (*Nasonia*
803 *vitripennis*). *Genes Genome Genetics*, *4*(March), 383–388. doi:10.1534/g3.113.008953

804 Beissbarth, T., & Speed, T. P. (2004). Gostat: find statistically overrepresented Gene Ontologies
805 within a group of genes. *Bioinformatics*, *20*(9), 1464–1465.
806 doi:10.1093/bioinformatics/bth088

807 Bewick, A. J., Ji, L., Niederhuth, C. E., Willing, E., Hofmeister, B. T., & Shi, X. (2016). On the
808 origin and evolutionary consequences of gene body DNA methylation. *PNAS*, *113*(32),
809 9111–9116. doi:10.1073/pnas.1604666113

810 Bewick, A. J., Vogel, K. J., Moore, A. J., & Schmitz, R. J. (2017). Evolution of DNA
811 methylation across insects. *Molecular Biology and Evolution*, *34*(3), 654–665.

812 doi:10.1093/molbev/msw264

813 Biergans, S. D., Jones, J. C., Treiber, N., Galizia, C. G., & Szyszka, P. (2012). DNA methylation
814 mediates the discriminatory power of associative long-term memory in honeybees. *PloS*
815 *One*, 7(6), 1–7. doi:10.1371/journal.pone.0039349

816 Bird, A. P. (1986). CpG-rich islands and the function of DNA methylation. *Nature*, 321, 209–
817 213.

818 Bird, A. P., & Wolffe, A. P. (1999). Methylation-Induced repression - Belts, braces, and
819 chromatin. *Cell*, 99, 451–454.

820 Bolger, A. M., Lohse, M., & Usadel, B. (2014). Genome analysis Trimmomatic : a flexible
821 trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120.
822 doi:10.1093/bioinformatics/btu170

823 Bonasio, R., Li, Q., Lian, J., Mutti, N. S., Jin, L., Zhao, H., ... Reinberg, D. (2012). Genome-
824 wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and
825 *Harpegnathos saltator*. *Current Biology*, 22(19), 1755–64. doi:10.1016/j.cub.2012.07.042

826 Busch, R., Qiu, W., Lasky-Su, J., Morrow, J., Criner, G., & DeMeo, D. (2016). Differential
827 DNA methylation marks and gene comethylation of COPD in African-Americans with
828 COPD exacerbations. *Respiratory Research*, 17(1), 1–15. doi:10.1186/s12931-016-0459-8

829 Cheng, X., & Blumenthal, R. M. (2008). Mammalian DNA methyltransferases: a structural
830 perspective. *Structure*, 16(3), 341–350.

831 Colot, V., & Rossignol, J. (1999). Eukaryotic DNA methylation as an evolutionary device.
832 *Bioessays*, 21, 402–411.

833 Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO:
834 a universal tool for annotation, visualization and analysis in functional genomics research.
835 *Bioinformatics*, 21(18), 3674–3676.

836 Davies, M. N., Volta, M., Pidsley, R., Lunnon, K., Dixit, A., Lovestone, S., ... Mill, J. (2012).
837 Functional annotation of the human brain methylome identifies tissue-specific epigenetic
838 variation across brain and blood. *Genome Biology*, 13(6), R43. doi:10.1186/gb-2012-13-6-
839 r43

840 Dhaygude, K., Nair, A., Johansson, H., Wurm, Y., & Sundström, L. (2018). The first draft
841 genomes of the ant *Formica exsecta*, and its *Wolbachia* endosymbiont reveal extensive gene
842 transfer from endosymbiont to host. *BioRxiv*. doi:http://dx.doi.org/10.1016/B978-

843 012168110-4/50000-X

844 Ecker, J. R., & Davis, R. W. (1986). Inhibition of gene expression in plant cells by expression of
845 antisense RNA. *PNAS*, 83(August), 5372–5376.

846 Eden, S., & Cedar, H. (1994). Role of DNA methylation in the regulation of transcription.
847 *Current Opinion in Genetics and Development*, 4, 255–259.

848 Eijk, K. R. Van, Jong, S. De, Boks, M. P. M., Langeveld, T., Colas, F., Veldink, J. H., ...
849 Ophoff, R. A. (2012). Genetic analysis of DNA methylation and gene expression levels in
850 whole blood of healthy human subjects. *BMC Genomics*, 13:636.

851 Elango, N., Hunt, B. G., Goodisman, M. a D., & Yi, S. V. (2009). DNA methylation is
852 widespread and associated with differential gene expression in castes of the honeybee, *Apis*
853 *mellifera*. *Proceedings of the National Academy of Sciences of the United States of*
854 *America*, 106(27), 11206–11. doi:10.1073/pnas.0900301106

855 Falckenhayn, C., Carneiro, V. C., Amarante, A. D. M., Schmid, K., Hanna, K., Kang, S., ...
856 Lyko, F. (2016). Comprehensive DNA methylation analysis of the *Aedes aegypti* genome.
857 *Scientific Reports*, 6, 36444. doi:10.1038/srep36444

858 Feng, S., Cokus, S. J., Zhang, X., Chen, P.-Y., Bostick, M., Goll, M. G., ... Jacobsen, S. E.
859 (2010). Conservation and divergence of methylation patterning in plants and animals.
860 *Proceedings of the National Academy of Sciences of the United States of America*, 107(19),
861 8689–8694. doi:10.1073/pnas.1002720107

862 Flores, K., Wolschin, F., Corneveaux, J., Allen, A., Huentelman, M., & Amdam, G. (2012).
863 Genome-wide association between DNA methylation and alternative splicing in an
864 invertebrate. *BMC Genomics*, 13(1), 480. doi:10.1186/1471-2164-13-480

865 Foret, S., Kucharski, R., Pellegrini, M., Feng, S., Jacobsen, S. E., Robinson, G. E., & Maleszka,
866 R. (2012). DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative
867 phenotypes in honey bees. *Proceedings of the National Academy of Sciences of the United*
868 *States of America*, 109(13), 4968–4973. doi:10.1073/pnas.1202392109

869 Glastad, K. M., Gokhale, K., Liebig, J., & Goodisman, M. A. D. (2016). The caste- and sex-
870 specific DNA methylome of the termite *Zootermopsis nevadensis*. *Scientific Reports*,
871 6(August), 1–14. doi:10.1038/srep37110

872 Glastad, K. M., Hunt, B. G., & Goodisman, M. A. D. (2012). Evidence of a conserved functional
873 role for DNA methylation in termites. *Insect Molecular Biology*, 22(2), 143–154.

874 doi:10.1111/imb.12010

875 Goll, M. G., & Bestor, T. H. (2005). Eukaryotic cytosine methyltransferases. *Annu. Rev.*
876 *Biochem.*, 74, 481–514.

877 Gu, H., Smith, Z. D., Bock, C., Boyle, P., Gnirke, A., & Meissner, A. (2011). Preparation of
878 reduced representation bisulfite sequencing libraries for genome-scale DNA methylation
879 profiling. *Nature Protocols*, 6(4), 468.

880 Hata, K., Okano, M., Lei, H., & Li, E. (2002). Dnmt3L cooperates with the Dnmt3 family of de
881 novo DNA methyltransferases to establish maternal imprints in mice. *Development*
882 *(Cambridge, England)*, 129(8), 1983–93. Retrieved from
883 <http://www.ncbi.nlm.nih.gov/pubmed/11934864>

884 He, X., & Zhang, J. (2006). Why do hubs tend to be essential in protein networks? *PLoS*
885 *Genetics*, 2(6), 0826–0834. doi:10.1371/journal.pgen.0020088

886 Herb, B. R., Shook, M. S., Fields, C. J., & Robinson, G. E. (2018). Defense against territorial
887 intrusion is associated with DNA methylation changes in the honey bee brain. *BMC*
888 *Genomics*, 19(1), 1–11. doi:10.1186/s12864-018-4594-0

889 Herb, B. R., Wolschin, F., Hansen, K. D., Aryee, M. J., Langmead, B., Irizarry, R., ... Feinberg,
890 A. P. (2012). Reversible switching between epigenetic states in honeybee behavioral
891 subcastes. *Nature Neuroscience*, 15(10), 1371–1373.

892 Horvath, S., Langfelder, P., Kwak, S., Aaronson, J., Rosinski, J., Vogt, T. F., ... Yang, X. W.
893 (2016). Huntington’s disease accelerates epigenetic aging of human brain and disrupts DNA
894 methylation levels. *Aging*, 8(7), 1485–1512. doi:10.18632/aging.101005

895 Horvath, S., Zhang, Y., Langfelder, P., Kahn, R. S., Boks, M. P. M., Eijk, K. Van, ... Ophoff, R.
896 A. (2012). Aging effects on DNA methylation modules in human brain and blood tissue.
897 *Genome Biology*, 13:R97.

898 Hunt, B. G., Glastad, K. M., Yi, S. V., & Goodisman, M. A. D. (2013a). Patterning and
899 regulatory associations of DNA methylation are mirrored by histone modifications in
900 insects. *Genome Biology and Evolution*, 5(3), 591–598. doi:10.1093/gbe/evt030

901 Hunt, B. G., Glastad, K. M., Yi, S. V., & Goodisman, M. A. D. (2013b). The function of
902 intragenic DNA methylation: insights from insect epigenomes. *Integrative and Comparative*
903 *Biology*, 53(2), 319–328.

904 Jaffe, A. E., Feinberg, A. P., Irizarry, R. A., & Leek, J. T. (2012). Significance analysis and

905 statistical dissection of variably methylated regions. *Biostatistics*, 13(1), 166–178.
906 doi:10.1093/biostatistics/kxr013

907 Jirtle, R. L., & Skinner, M. K. (2007). Environmental epigenomics and disease susceptibility.
908 *Nature Reviews Genetics*, 8(April), 253–262. doi:10.1038/nrg2045

909 Kato, Y., Kaneda, M., Hata, K., Kumaki, K., Hisano, M., Kohara, Y., ... Sasaki, H. (2007). Role
910 of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during
911 male germ cell development in the mouse. *Human Molecular Genetics*, 16(19), 2272–80.
912 doi:10.1093/hmg/ddm179

913 Kay, S., Skowronski, D., & Hunt, B. G. (2017). Developmental DNA methyltransferase
914 expression in the fire ant *Solenopsis invicta*. *Insect Science*, 00, 1–9. doi:10.1111/1744-
915 7917.12413

916 Keshet, I., Yisraeli, J., & Cedar, H. (1985). Effect of regional DNA methylation on gene
917 expression. *PNAS*, 82(May), 2560–2564.

918 Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2 :
919 accurate alignment of transcriptomes in the presence of insertions, deletions and gene
920 fusions. *Genome Biology*, 14, R36.

921 Kim, J., Samaranyake, M., & Pradhan, S. (2009). Epigenetic mechanisms in mammals. *Cellular
922 and Molecular Life Sciences*, 66(4), 596.

923 Kozeretska, I. A., Serga, S. V, Koliada, A. K., & Vaiserman, A. M. (2017). *Epigenetic
924 Regulation of Longevity in Insects*. *Insect Epigenetics* (1st ed., Vol. 53). Elsevier Ltd.
925 doi:10.1016/bs.aiip.2017.03.001

926 Krueger, F., & Andrews, S. R. (2011). Bismark: a flexible aligner and methylation caller for
927 Bisulfite-Seq applications. *Bioinformatics*, 27(11), 1571–1572.
928 doi:10.1093/bioinformatics/btr167

929 Kucharski, R., Maleszka, J., Foret, S., & Maleszka, R. (2008). Nutritional control of reproductive
930 status in honeybees via DNA methylation. *Science*, 319(5871), 1827–1830.

931 Langfelder, P., & Horvath, S. (2007). Eigengene networks for studying the relationships between
932 co-expression modules. *BMC Systems Biology*, 1(1), 54.

933 Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network
934 analysis. *BMC Bioinformatics*, 9, 559. doi:10.1186/1471-2105-9-559

935 Langfelder, P., Luo, R., Oldham, M. C., & Horvath, S. (2011). Is my network module preserved

936 and reproducible? *PLoS Computational Biology*, 7, e1001057.
937 doi:10.1371/journal.pcbi.1001057

938 Langfelder, P., Mischel, P. S., & Horvath, S. (2013). When Is Hub Gene Selection Better than
939 Standard Meta-Analysis? *PLoS ONE*, 8(4). doi:10.1371/journal.pone.0061505

940 Li-Byarlay, H., Li, Y., Stroud, H., Feng, S., Newman, T. C., Kaneda, M., ... Robinson, G. E.
941 (2013). RNA interference knockdown of DNA methyltransferase 3 affects gene alternative
942 splicing in the honey bee. *PNAS*, 110(31), 12750–12755. doi:10.1073/pnas.1310735110/
943 /DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1310735110

944 Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data
945 with or without a reference genome. *BMC Bioinformatics*, 12, 323. doi:10.1186/1471-2105-
946 12-323

947 Li, E., Beard, C., & Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting.
948 *Nature*, 366, 362.

949 Libbrecht, R., Oxley, P. R., Keller, L., Jan, D., & Kronauer, C. (2016). Robust DNA Methylation
950 in the Clonal Raider Ant. *Current Biology*, 26(3), 391–395. doi:10.1016/j.cub.2015.12.040

951 Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Nery, J. R., ... Ecker, J. R.
952 (2009). Human DNA methylomes at base resolution show widespread epigenomic
953 differences. *Nature*, 462(7271), 315–322. doi:10.1038/nature08514.Human

954 Liu, L., Li, Y., & Tollefsbol, T. O. (2008). Gene-environment interactions and epigenetic basis
955 of human diseases. *Current Issues in Molecular Biology*, 10(1–2), 25.

956 Lockett, G. A., Helliwell, P., & Maleszka, R. (2010). Involvement of DNA methylation in
957 memory processing in the honey bee. *Neuroreport*, 21(12), 812–816.

958 Lyko, F., Foret, S., Kucharski, R., Wolf, S., Falckenhayn, C., & Maleszka, R. (2010). The honey
959 bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS*
960 *Biology*, 8(11), e1000506. doi:10.1371/journal.pbio.1000506

961 Lyko, F., & Maleszka, R. (2011). Insects as innovative models for functional studies of DNA
962 methylation. *Trends in Genetics*, 27(4), 127–131.

963 Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads.
964 *EMBnet J*, 17(17), 10–12.

965 Maunakea, A. K., Nagarajan, R. P., Bilenky, M., Ballinger, T. J., Souza, D., Fouse, S. D., ...
966 Costello, J. F. (2010). Conserved role of intragenic DNA methylation in regulating

967 alternative promoters. *Nature*, 466(7303), 253–257. doi:10.1038/nature09165. Conserved
968 Morandin, C., Dhaygude, K., Paviola, J., Trontti, K., Wheat, C., & Helanterä, H. (2015). Caste-
969 biases in gene expression are specific to developmental stage in the ant *Formica exsecta*.
970 *Journal of Evolutionary Biology*, 28(9), 1705–1718. doi:10.1111/jeb.12691
971 Morandin, C., Tin, M. M. Y., Abril, S., Gómez, C., Pontieri, L., Schiøtt, M., ... Mikheyev, A. S.
972 (2016). Comparative transcriptomics reveals the conserved building blocks involved in
973 parallel evolution of diverse phenotypic traits in ants. *Genome Biology*, 1–19.
974 doi:10.1186/s13059-016-0902-7
975 Okano, M., Bell, D. W., Haber, D. A., & Li, E. (1999). DNA methyltransferases Dnmt3a and
976 Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, 99(3),
977 247–257.
978 Okano, M., Xie, S., & Li, E. (1998). Cloning and characterization of a family of novel
979 mammalian DNA (cytosine-5) methyltransferases. *Nature Genetics*, 19(3), 219.
980 Page, R. E., & Peng, C. Y. (2001). Aging and development in social insects with emphasis on the
981 honey bee, *Apis mellifera* L. *Experimental Gerontology*, 36(4–6), 695–711. Retrieved from
982 <http://www.ncbi.nlm.nih.gov/pubmed/11295509>
983 Pamilo, P. (1991). Life span of queens in the ant *Formica exsecta*. *Insectes Sociaux*, 38(2), 111–
984 119.
985 Patalano, S., Vlasova, A., Wyatt, C., Ewels, P., Camara, F., & Ferreira, P. G. (2015). Molecular
986 signatures of plastic phenotypes in two eusocial insect species with simple societies. *PNAS*,
987 112(45), 13970–13975. doi:10.1073/pnas.1515937112
988 Razin, a, & Cedar, H. (1991). DNA methylation and gene expression. *Microbiological Reviews*,
989 55(3), 451–8. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15881894>
990 Richardson, B. (2003). Impact of aging on DNA methylation. *Ageing Research Reviews*, 2(3),
991 245–261. doi:10.1016/S1568-1637(03)00010-2
992 Rickabaugh, T. M., Baxter, R. M., Sehl, M., Sinsheimer, J. S., Hultin, P. M., Hultin, L. E., ...
993 Jamieson, B. D. (2015). Acceleration of age-associated methylation patterns in HIV-1-
994 infected adults. *PLoS ONE*, 10(3), 1–18. doi:10.1371/journal.pone.0119201
995 Robinson, G. E., Grozinger, C. M., & Whitfield, C. W. (2005). Sociogenomics: social life in
996 molecular terms. *Nature Reviews. Genetics*, 6(4), 257–70. doi:10.1038/nrg1575
997 Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for

998 differential expression analysis of digital gene expression data. *Bioinformatics (Oxford,*
999 *England)*, 26(1), 139–40. doi:10.1093/bioinformatics/btp616

1000 Schübeler, D. (2015). Function and information content of DNA methylation. *Nature*, 517, 321–
1001 326. doi:10.1038/nature14192

1002 Shen, L., & Sinai, M. (2016). GeneOverlap: test and visualize gene overlaps. R package version
1003 1.2.0. 2013.

1004 Shi, Y. Y., Yan, W. Y., & Huang, Z. Y. (2013). Genomewide analysis indicates that queen larvae
1005 have lower methylation levels in the honey bee (*Apis mellifera*). *Naturwissenschaften*,
1006 100, 193–197. doi:10.1007/s00114-012-1004-3

1007 Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., & Kashlev, M. (2011).
1008 CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature*,
1009 479(7371), 74–79. doi:10.1038/nature10442

1010 Smith, C. R., Mutti, N. S., Jasper, W. C., Naidu, A., Smith, C. D., & Gadau, J. (2012). Patterns
1011 of DNA methylation in development, division of labor and hybridization in an ant with
1012 genetic caste determination. *PloS One*, 7(8), e42433. doi:10.1371/journal.pone.0042433

1013 Standage, D. S., Berens, A. L. I. J., Glastad, K. M., & Severin, A. J. (2016). Genome ,
1014 transcriptome and methylome sequencing of a primitively eusocial wasp reveal a greatly
1015 reduced DNA methylation system in a social insect. *Molecular Ecology*, 25, 1769–1784.
1016 doi:10.1111/mec.13578

1017 Suzuki, M. M., & Bird, A. (2008). DNA methylation landscapes: provocative insights from
1018 epigenomics. *Nature Reviews Genetics*, 9(June), 465–476. doi:10.1038/nrg2341

1019 Tan, J. A., & Mikheyev, A. S. (2016). A scaled-down workflow for Illumina shotgun sequencing
1020 library preparation : lower input and improved performance at small fraction of the cost.
1021 *PeerJ Preprints*.

1022 Tate, P. H., & Bird, P. (1993). Effects of DNA methylation on proteins and gene expression.
1023 *Current Opinion in Genetics and Development*, 3, 226–231.

1024 Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., ... Pachter, L. (2012).
1025 Differential gene and transcript expression analysis of RNA-seq experiments with TopHat
1026 and Cufflinks. *Nature Protocols*, 7(3), 562–578. doi:10.1038/nprot.2012.016

1027 Vilcinskas, A. (2016). The role of epigenetics in host – parasite coevolution : lessons from the
1028 model host insects *Galleria mellonella* and *Tribolium castaneum* □. *Zoology*, 119(4), 273–

1029 280. doi:10.1016/j.zool.2016.05.004

1030 Wang, Y., Jorda, M., Jones, P. L., Maleszka, R., Ling, X., Robertson, H. M., ... Robinson, G. E.
1031 (2007). Functional CpG Methylation System in a Social Insect. *Science*, 314, 645–647.
1032 doi:10.1126/science.1229223

1033 Waterland, R. A., & Jirtle, R. L. (2003). Transposable Elements : Targets for Early Nutritional
1034 Effects on Epigenetic Gene Regulation. *Molecular and Cellular Biology*, 23(15), 5293–
1035 5300. doi:10.1128/MCB.23.15.5293

1036 Weber, M., Hellmann, I., Stadler, M. B., Ramos, L., Pääbo, S., Rebhan, M., & Schübeler, D.
1037 (2007). Distribution , silencing potential and evolutionary impact of promoter DNA
1038 methylation in the human genome. *Nature Genetics*, 39(4), 457–466. doi:10.1038/ng1990

1039 Wedd, L., Kucharski, R., & Maleszka, R. (2016). Differentially methylated obligatory epialleles
1040 modulate context-dependent LAM gene expression in the honeybee *Apis mellifera*.
1041 *Epigenetics*, 11(1), 1–10. doi:10.1080/15592294.2015.1107695

1042 Wolffe, A. P., & Matzke, M. A. (1999). Epigenetics : Regulation Through Repression. *Science*,
1043 286(481), 481–486. doi:10.1126/science.286.5439.481

1044 Xiang, H., Zhu, J., Chen, Q., Dai, F., Li, X., Li, M., ... Wang, J. (2010). Single base-resolution
1045 methylome of the silkworm reveals a sparse epigenomic map. *Nature Biotechnology*, 28(5),
1046 516–520. doi:10.1038/nbt.1626

1047 Yan, H., Bonasio, R., Simola, D. F., & Berger, S. L. (2015). DNA methylation in social insects :
1048 how epigenetics can control behavior and longevity. *Annual Review of Entomology*, 60(60),
1049 435–452. doi:10.1146/annurev-ento-010814-020803

1050 Yang, S., Guo, C., Zhao, X., Sun, J., & Hong, X. (2017). Divergent methylation pattern in adult
1051 stage between two forms of *Tetranychus urticae* (Acari : Tetranychidae). *Insect Science*,
1052 00, 1–12. doi:10.1111/1744-7917.12444

1053 Zemach, A., Mcdaniel, I. E., Silva, P., & Zilberman, D. (2010). Genome-wide evolutionary
1054 analysis of eukaryotic DNA methylation. *Science*, 328(April), 916–919.

1055 Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression network
1056 analysis. *Statistical Applications in Genetics and Molecular Biology*, 4(1), 17.

1057 Zhang, J., & Huang, K. (2017). Pan-cancer analysis of frequent DNA co- methylation patterns
1058 reveals consistent epigenetic landscape changes in multiple cancers. *BMC Genomics*,
1059 18(Suppl 1), 1–14. doi:10.1186/s12864-016-3259-0

1060 Zilberman, D., Coleman-Derr, D., Ballinger, T., & Henikoff, S. (2008). Histone H2A.Z and
1061 DNA methylation are mutually antagonistic chromatin marks. *Nature*, 456(7218), 125–129.

1062

1063

1064 **Data Accessibility Statement**

1065 The raw reads are publicly available in the DNA Data Bank of Japan under bioproject ID

1066 PRJDB6378. The raw reads of the transcriptome are publicly available under sample accession

1067 numbers ID SAMD00094824 - SAMD00094838. The raw reads of the methylome are publicly

1068 available under sample accession numbers ID SAMD00094839 - SAMD00094858.

1069

1070 **Authors contribution**

1071 Conceptualization, CM; Methodology, CM and ASM; Fieldwork: HH and LS; Investigation,

1072 CM, ASM and VPB; Writing, CM and ASM; Review & Editing, All authors; Funding

1073 Acquisition, CM, HH, LS, ASM. All authors read and approved the final manuscript.

1074

1075 **Competing interests**

1076 The authors declare that they have no competing interests.

1077

1078 **Figure legends**

1079 **Figure 1**

1080 **Mapping of methylation sites on the genome annotation.** Output was generated by

1081 `BWASPR::map_methylome()` and accounts for where CpG_{hsm} and CpG_{scd} (control) sites reside

1082 relative to the genome annotation, in every sample. a) All regions b) Exon regions. Cytosine

1083 DNA methylation was found almost exclusively in a CpG context. Highly supported DNA

1084 methylated CpGs sites are observed at ~2.5-fold higher numbers in coding regions than expected
1085 by random distribution over every scd site.

1086

1087 **Figure 2**

1088 **Scatter plot showing the correlation between expression level (FPKM) and average**

1089 **percentage of DNA methylation per site.** A positive correlation between expression and DNA

1090 methylation was found, implying that highly expressed genes are more likely to be highly

1091 methylated ($cor = 0.28$, $p < 0.001$; Spearman's product-moment correlation). This is the typical

1092 pattern found in insect genomes that rely on DNA methylation (Hunt, Glastad, Yi, &

1093 Goodisman, 2013b; Xiang et al., 2010), providing a level of validation for these findings.

1094

1095 **Figure 3**

1096 **Venn diagram summarizing overlap between differentially expressed and differentially**

1097 **methylated genes between all comparisons at the gene level.** No significant overlap was found

1098 at the gene level for any of the four comparisons. Statistical analysis of the significance of the

1099 overlap between the two gene sets was calculated using the GeneOverlap BioConductor package.

1100

1101 **Figure 4**

1102 **Venn diagram summarizing overlap between GO terms of differentially expressed and**

1103 **differentially methylated genes between comparisons.** Significant overlap was found at the

1104 GO term level for all comparisons. Statistical analysis of the significance of the overlap between

1105 the two gene sets was calculated using the GeneOverlap BioConductor package.

1106

1107 **Figure 5**

1108 **Visual representation of the link between co-expression and co-methylation modules.**

1109 Many modules belonging to the co-expression and co-methylation networks were correlated with
1110 either caste (inner band), or stage (outer bands). **Significant correlations with the queen caste**
1111 **and developmental maturity are highlighted in red, while worker and newly eclosed**
1112 **developmental stage correlations are highlighted in blue. Significant caste by**
1113 **developmental stage interactions are highlighted in purple.** Although conservation of
1114 modules between co-expression and co-methylation datasets was generally weak, four
1115 expression modules (E1, E6, E9 and E10) were conserved in the co-methylation data. Their
1116 connections are highlighted in orange.

1117

1118 **Figure 6**

1119 **Preservation and association of co-expression and co-methylation modules using the**

1120 **Module preservation statistic, $Z_{summary}$, of WGCNA.** a) preservation of expression modules
1121 in the methylation dataset b) preservation of methylation modules in the expression dataset.

1122 Values of $Z_{summary}$ below 2 indicate no evidence of preservation; values between 2 and 5
1123 indicate moderate evidence for preservation; values above 5 strong evidence for preservation.

1124 Four co-expression modules (E1, E6, E9 and E10) were conserved in the co-methylation data.

1125

1126 **Table legends**

1127 **Table 1**

1128 **Correlation between module eigengenes and biological traits (caste and stage, and**

1129 **caste/stage interactions).** Expression and DNA methylation patterns of most modules are

1130 strongly associated with developmental stage phenotype. In addition, expression of several of
1131 these modules was also associated for both phenotypes and phenotype interactions. This shows
1132 that modules likely play multiple roles, and that their constituent genes have many functions.
1133 Caste/black: module associated with queen phenotype. Caste/grey: module associated with
1134 worker phenotype. Stage/black: module associated with old phenotype. Stage/grey: module
1135 associated with emerging phenotype. Caste x Stage/grey/black: module associated with both
1136 caste and stage.

1137

1138 **Supplementary figures**

1139 **Figure S1**

1140 **Histogram of CpG read coverage per cytosine for all aggregate samples of emerging queen,**
1141 **old queen, emerging worker, old worker.** The range of coverage was restricted to the range
1142 [10-1000] for easy comparison. Plots were generated with methylKit:getCoverageStats() as
1143 implemented in the BWASPR workflow.

1144

1145 **Figure S2**

1146 **Histogram of % DNA methylation per CpG site for all aggregate samples of emerging**
1147 **queens, old queens, emerging workers, old workers.** The range of coverage was restricted to
1148 the range [10-1000] for easy comparison. Plots were generated with
1149 methylKit:getMethylationStats() as implemented in the BWASPR workflow.

1150

1151 **Figure S3**

1152 **Barplot showing the number of genes differentially expressed and differentially methylated**
1153 **across castes (Old queens (OQ) vs. old Workers (OW); Emerging queens (EQ) vs. Emerging**

1154 Workers (EW)) and across stages (Emerging Queens (EQ) vs. Old Queens (OQ); Emerging
1155 Workers (EW) vs. Old Workers (OW)). A larger number of genes were found differentially
1156 expressed than differentially methylated. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

1157

1158 **Figure S4**

1159 **Venn diagram summarizing numbers of genes that are found always differentially**
1160 **expressed/methylated between castes regardless of the developmental stage (OQW vs.**
1161 **EQW) and genes that are always differentially expressed/methylated between**
1162 **developmental stages regardless of the caste (Queen vs. Worker). Three of the four**
1163 **comparisons presented a significant overlap, implying that a larger number of genes is**
1164 **consistently differentially expressed by caste or developmental stage than by chance. For**
1165 **methylation, a larger number of genes was consistently caste differentially methylated across**
1166 **stages than expected by chance. Statistical analysis of the significance of the overlaps between**
1167 **the two gene sets was calculated using the GeneOverlap BioConductor package.**

1168

1169 **Figure S5**

1170 **Differentially expressed and differentially methylated genes across comparisons (Old**
1171 **queens vs. Old workers; Emerging queens vs. Emerging workers) visualized as ‘MA’ plots**
1172 **(log ratio versus abundance).**

1173

1174 **Figure S6**

1175 **Caste- and developmental stage- bias direction of the genes that were both differentially**
1176 **expressed and differentially methylated across all treatments (OQ vs. OW; EQ vs. EW; EQ**

1177 vs. **OQ; EW vs. OW**). Only half of the genes that were over-expressed in one direction tended to
1178 be also over-expressed in the same direction, apart from the emerging queens vs. emerging
1179 workers comparison, where all eight genes were over-expressed and over-methylated in
1180 emerging queens.

1181

1182 **Figure S7**

1183 **Preservation and association of *F. exsecta* brain co-expression and 16 ant species co-**
1184 **expression network published in an earlier study (Morandin et al. 2016) using the Module**
1185 **preservation statistic *Zsummary* of WGCNA.** a) preservation of expression modules in the
1186 2016 expression data b) preservation of expression modules in the 2016 expression data. Values
1187 of “*Zsummary* below 2” indicate no evidence of preservation. Values between 2 and 5 indicate
1188 moderate evidence for preservation. Values above 5 demonstrate strong evidence for
1189 preservation. Three 2016 expression modules were preserved in the present study expression
1190 modules, while 9 expression modules from the present study were preserved in the 2016
1191 expression modules.

1192

1193 **Supplementary tables**

1194 **Table S1**

1195 **RRBS primer sequences.** Sequences of RRBS primer used for the RRBS library preparation
1196 (see methods for more details).

1197

1198 **Table S2**

1199 **Statistics of deposited RRBS and RNAseq samples.** All methylation data were derived with
1200 the BWASP/BWASPR workflows. Median length was determined with FastQC. Mapping
1201 efficiency was determined with Bismark software, as implemented in BWASPR. All expression
1202 data were derived with the Tophat2, cufflinks and RSEM pipeline.

1203

1204 **Table S3**

1205 **Details of the analysis performed to find minimal coverage of sites required to potentially**
1206 **observe significant DNA methylation.** Following this analysis, a cut-off of 4 reads was set as
1207 the threshold to detect sites with enough coverage (scd sites) to perform detection of significant
1208 DNA methylation.

1209

1210 **Table S4**

1211 **Dataframe used as input for the co-methylation network analysis with WGCNA.** Each
1212 column represents one sample analyzed and each row represents the DNA methylation level of
1213 one gene across all samples. The input dataset consists of results from DNA methylation calls of
1214 the 12,112 genes measured, as the average percentage of DNA methylation per site per gene,
1215 restricted to sites with highly-supported DNA methylation data (high coverage, at least 10 reads).
1216 These proportional levels control for the number of restriction enzyme sites present on each
1217 gene, gene length, and quality of mapping to avoid any biases.

1218

1219 **Table S5**

1220 **Dataframe used as input for the co-expression network analysis with WGCNA.** Each
1221 column represents one sample analyzed and each row represents the expression level of one gene

1222 across all samples. The input dataset consists of the expression level (raw counts) of the 13,041
1223 genes evaluated. Raw counts were log-transformed as recommended by Langfelder & Horvath
1224 (2008).

1225

1226 **Table S6**

1227 **Overall level of DNA methylation** (percent reads reporting conversion or non-conversion) on
1228 all C, CpG sites, CHG sites and CpG sites for all biological samples (old queens, emerging
1229 queens, old workers, emerging workers).

1230

1231 **Table S7**

1232 **Number of CpG sites, CHG sites, CHH sites, for each biological sample (old queens,
1233 emerging queens, old workers, emerging workers).**

1234

1235 **Table S8**

1236 **Per aggregate sample DNA methylation statistics.** scd, sufficiently covered detectable sites,
1237 and hsm, highly supported methylated sites, were reported by the BWASPR workflow. The scd
1238 percentage refers to the proportion of genomic CpG that are scd. The hsm percentage refers to
1239 the proportion of scd that are hsm. The overall DNA methylation level refers to the proportion of
1240 Cs in all mapped reads that are called methylated, as determined by Bismark software as
1241 implemented in the BWASPR workflow.

1242

1243 **Table S9**

1244 **Distribution of CpG (hsm) sites in genome feature regions** for all biological samples (old
1245 queens, emerging queens, old workers, emerging workers). Cytosine DNA methylation was
1246 found almost exclusively in CpG contexts, as expected for insects.

1247

1248 **Table S10**

1249 **Distribution of CpG (hsm) sites in exon feature regions** for all biological samples (old queens,
1250 emerging queens, old workers, emerging workers). Highly supported DNA methylated CpGs
1251 sites are observed at about 2.5-fold higher frequencies in coding regions than expected by
1252 random distribution over every scd site.

1253

1254 **Table S11**

1255 **List of differentially methylated sites** as determined by methylKit::getMethylDiff with qvalue
1256 < 0.01. The list of differentially methylated CpG regions was based on q-values (0.01) and
1257 percent DNA methylation difference cutoffs (25%) for sites with coverage of at least 10 reads,
1258 for all four comparisons using for multiple testing correction.

1259

1260 **Table S12**

1261 **List of differentially expressed and differentially methylated genes for all four comparisons**
1262 **(OQ vs. OW; EQ vs. EW, OQ vs. EQ; OW vs. EW). Blast annotations for the genes are**
1263 **provided, as well as differential analysis results (EdgeR results for DEG and BWASP**
1264 **outputs for DMG).** The software Blast2GO (www.blast2go.com) was used to infer functional
1265 annotation of the *F. exsecta* gene set using structural similarity (BLASTx with an e-value cut-off
1266 $\leq 10^{-3}$).

1267

1268 **Table S13**

1269 **List of enriched GO terms for differentially expressed and differentially methylated genes**

1270 **found in all four comparisons.** Gene ontology (GO) terms for all genes of the *F. exsecta* gene

1271 set were determined using BLAST2GO (using BLASTp with an e-value cut-off $\leq 10^{-3}$. We used

1272 the GOstats package in R to conduct GO term enrichment analysis on the list of differentially

1273 expressed/methylated genes presented in Table S12, using the set of all genes for which GO

1274 terms were available, as the universe.

1275

1276 **Table S14**

1277 **List of enriched GO term for each module found in the co-methylation and co-expression**

1278 **network.** We used the GOstats package for R to conduct GO term enrichment analysis on gene

1279 sets included in the co-expression and co-methylation modules, using the set of all genes for

1280 which GO terms were available as the universe. Modules were enriched with gene ontology

1281 categories, which provides indirect evidence that these sets of co-expressed genes are

1282 biologically meaningful.

1283

1284 **Table S15**

1285 **List of blast annotations for all genes belonging to a co-methylation or a co-expression**

1286 **module.** The software Blast2GO (www.blast2go.com) was used to infer functional annotation of

1287 the *F. exsecta* gene set using structural similarity (BLASTx with an e-value cut-off $\leq 10^{-3}$).

1288

1289 **Table S16**

1290 **Preservation between co-expression and co-methylation networks.** Module preservation
1291 statistic *Zsummary* that summarizes evidence of preservation of expression modules in DNA
1292 methylation data, and preservation of DNA methylation modules in expression data. Values of
1293 *Zsummary* below 2 indicate no evidence of preservation; values between 2 and 5 indicate
1294 moderate evidence for preservation; values above 5 strong evidence for preservation. Four co-
1295 expression modules (E1, E6, E9, and E10) were conserved in the co-methylation data.

1296

1297 **Table S17**

1298 **Preservation between the brain co-expression network and the co-expression network from**
1299 **Morandin et al. (2016).** The module preservation statistic, *Zsummary*, that summarizes evidence
1300 of preservation of expression modules in our 2016 study (Morandin et al. 2016), and vice versa.
1301 Values of *Zsummary* below 2 indicate no evidence of preservation. Values between 2 and 5
1302 indicate moderate evidence for preservation. Values above 5 strong evidence for preservation.
1303 Three 2016 expression modules were preserved in the present study expression modules, while 9
1304 expression modules from the present study were preserved in the 2016 expression modules.

1305

1306 **Table S18**

1307 **List of genes defined as hub genes** for co-expression and co-methylation networks (High
1308 connectivity (($\text{cor.geneModuleMembership} > 0.8$) and high gene significance
1309 ($\text{cor.geneTraitSignificance} > 0.5$)). Hub genes are genes that are highly connected within a
1310 module, and that participate in biological processes associated with the modules (He & Zhang,
1311 2006; Langfelder & Horvath, 2008; Langfelder et al., 2013).

1312

1313 **Table S19**

1314 Expression and DNA methylation profiles of the ten hub genes belonging to module E6
1315 (correlated with stage, the interaction of caste x stage, and is preserved across the co-methylation
1316 modules). Two of the ten hub genes were also found differentially expressed across caste (OQ
1317 vs. OW) and across stages (OQ vs. EQ).

1318