bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Methylation and Gene Expression Differences Between Reproductive Castes of Bumblebee Workers.

Hollie Marshall[1], Zoë N. Lonsdale[1] and Eamonn B. Mallon[1]

[1]Department of Genetics and Genome Biology, The University of Leicester, Leicester, UK.

Thursday 10th January, 2019

***Keywords*—** methylation, bumblebee, expression, Hymenoptera, Bombus

Abstract

Phenotypic plasticity is the production of multiple phenotypes from a single genome and is notably observed in social insects. Multiple epigenetic mechanisms have been associated with social insect plasticity, with methylation being explored to the greatest extent. Methylation is known to play a role in caste determination in Apis mellifera but there is limited knowledge on it's role in other bee species. In this study we analysed whole genome bisulfite sequencing and RNA-seq data sets from head tissue of reproductive and sterile castes of the eusocial bumblebee Bombus terrestris. We found CpG methylation is enriched in coding regions and exons whereas non-CpG methylation is more evenly spread throughout the genome. Methylation has a positive relationship with gene expression and is found in lower levels in differentially expressed genes. We found no evidence for a role of methylation in alternative splicing. We did however find methylation differences between reproductive castes with some differentially methylated genes involved in behaviour and reproductive processes. Our results also indicate high inter-colony variation in methylation. This study provides the first insights into the nature of a bumblebee caste specific methylome as well as it's interaction with gene expression, alternative splicing and caste determination, providing greater understanding of the role of methylation in social insect phenotypic plasticity.

DRAFT

bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Phenotypic plasticity is the production of multiple phenotypes from a single genome. It plays a crucial role in the adaptive capabilities of species (Chevin et al., 2010) and is notably observed in social insects. Social insects exhibit, sometimes extreme, morphological and behavioural differences within a single colony known as castes. The mechanisms by which species develop differences between castes are diverse; some species use only environmental queues whilst others rely only on inherited changes, with many species falling somewhere in between these two extremes (Matsuura et al., 2018). For example some ant species from the *Pogonomyrmex* genus have purely genetic caste determination (Mott et al., 2015). On the other hand, many ant species undergo caste determination in response to only the environment, indicating their genomes must contain the code for all caste possibilities, with the phenotype potentially determined by epigenetic factors (Bonasio et al., 2012).

Multiple epigenetic mechanisms have been associated with social insect plasticity. Histone modifications have been shown to be involved with plasticity, for example changes in histone acetylation alter the behaviour of major workers of the ant species *Camponotus floridanus*, making them more similar to the behaviour of minor workers (Simola et al., 2016). Variation in microRNA expression levels has been identified to play a role in both honeybee (Ashby et al., 2016) and bumblebee (Collins et al., 2017) caste determination. However the most active research in this area has been focused on DNA methylation (Glastad et al., 2015).

DNA methylation is the addition of a methyl group to a cytosine nucleotide. In mammals the percentage of CpG's methylated is usually over 70%, with methylation serving to repress gene expression by occurring in promoter regions (Feng et al., 2010). However in insects is it generally found in considerably lower quantities, usually less than 1% of CpG's have methylation (Lyko et al., 2010; Wang et al., 2013), however see Glastad et al. 2016. It is also found in gene bodies rather than promoter regions (Fang et al., 2012; Wang et al., 2013), with a possible role in alternative splicing (Bonasio et al., 2012) rather than gene silencing as in mammals.

bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

DNA methylation has been associated with the switching of worker castes in honeybees (Herb et al., 2012). A major finding was that silencing of the *DNMT3* gene (involved in methylation establishment) in larvae produced queens rather than workers (Kucharski et al., 2008). DNA methylation has also been linked with alternative splicing differences between castes in two ant species (Bonasio et al., 2012) and thought to be involved in caste determination for a species of social wasp (Shaham et al., 2016).

However it is clear DNA methylation is not a conserved mechanism in Hymenoptera for caste differentiation. No association between caste and methylation has been found in a number of wasp and ant species (Standage et al., 2016; Patalano et al., 2015). Additionally, the statistical methods of previous next generation sequencing analyses on social insect methylation have been brought into question (Libbrecht et al., 2016).

Here we assess whole genome methylation differences between reproductive castes of the bumblebee, *Bombus terrestris*. Bumblebees are primitively eusocial and are an important pollinator species, both economically and environmentally. They are generalist pollinators and are keystone species in some ecosystems (Woodard et al., 2015). The *B.terrestris* genome has been sequenced with a relatively small genome size of around 249Mb (Sadd et al., 2015). Multiple recent studies have highlighted *B.terrestris* as an ideal organism to assess methylation as a potential regulatory mechanism for caste determination (Li et al., 2018; Lonsdale et al., 2017; Amarasinghe et al., 2014)

Methylation regulatory genes were identified in the bumblebee genome and have since been shown to have varying expression levels between queens, workers and drones (Li et al., 2018). Additionally genes showing allele-specific methylation and gene expression have been identified and are enriched in reproductive related processes (Lonsdale et al., 2017). Experimental changes in methylation in *B.terrestris* workers has been shown to alter levels of reproductive behaviour (Amarasinghe et al., 2014). Whilst these studies highlight differences in methylation between *B.terrestris* castes it is still unclear where those differences are within the genome and also whether methylation differences are related to changes in gene expression, potentially leading to caste

bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

differentiation.

In this study we analysed whole genome bisulfite sequencing datasets from reproductive and sterile worker castes of *B.terrestris* allowing us to identify base-pair resolution methylation throughout the genome. We then linked these data with gene expression data for the same individuals to identify a potential relationship between gene expression and methylation. We asked is methylation enriched in the coding regions of genes, as seen in other social insects (Patalano et al., 2015; Glastad et al., 2016; Bonasio et al., 2012). We also searched for a role of methylation in alternative splicing. Finally we checked if any differentially methylated genes identified were homologous to those found between honeybee reproductive castes, from previously published work.

bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

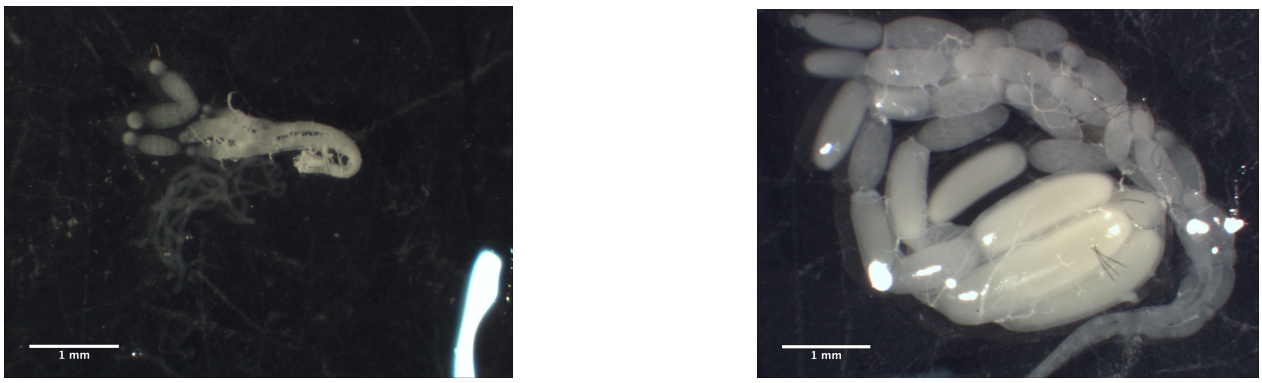
Methods

Bee husbandry and tissue sampling

Three *B. terrestris* colonies, from Agralan, UK, were reared in constant red light at 26°C and 60% humidity. They were fed 50% v/v apiary solution (Meliose-Roquette, France) and pollen (Percie du set, France) *ad libitum*. Callow workers, less than 24 hours old, were taken from each colony and placed in small rearing boxes of five individuals.

The worker bees were sacrificed at six days old. For each bee, the head was snap frozen in liquid nitrogen. Through dissection in 1% PBS solution, the reproductive status of each bee was determined and classed as either reproductive, sterile, or intermediate. Workers were classed as having developed ovaries, and therefore reproductive, if the largest oocyte was larger than the trophocyte follicle (Duchateau and Velthuis, 1988). This measurement is tightly correlated with reproductive status (Geva et al., 2005; Foster et al., 2004). The ovaries of each worker were weighed, and the length of the largest oocyte was measured using ImageJ v.1.50e (Schneider et al., 2012) (supplementary 1.0). Worker 'reproductiveness' was classified on a scale from 0-4 based on Duchateau and Velthuis (1988), 0 begin completely sterile (Fig.1a) and 4 having fully developed ovaries (Fig.1b).

bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



(a) Sterile Worker Ovary

(b) Reproductive Worker Ovary

Figure 1: (a) one half of a pair of ovaries from a sterile bumblebee worker, with a score of 0. (b) one half of a pair of ovaries from a reproductive bumblebee worker, with a score of 4. Scores generated following Duchateau and Velthuis (1988).

RNA and DNA extraction and sequencing

Bumblebee heads from reproductive and sterile workers were cut in half (using a lateral incision central between the eyes). Each head half was randomly allocated for either DNA/RNA extraction to avoid left/right hemisphere bias. RNA was extracted using the Sigma-Aldrich GenElute Mammalian Total RNA Miniprep kit and DNA was extracted using the Qiagen DNeasy blood and Tissue kit, following manufacturers protocols. The extracted RNA was treated with DNase and the extracted DNA was treated with RNase. DNA samples from the same colony and with the same reproductive status (reproductive or sterile) were pooled producing six pooled samples in total. DNA and RNA quality and quantity was determined by Nanodrop and Qubit® fluorometers. A total of 18 RNA samples were sent for 100bp paired-end sequencing and 6 pooled DNA samples were sent for 100bp paired-end bisulfite sequencing on a HiSeq 2000 machine (Illumina, Inc.) by BGI Tech Solution Co., Ltd.(Hong Kong).

bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Differential expression and alternative splicing

Low quality bases were removed from the RNA-Seq libraries using CutAdapt v1.1 (Martin, 2011). Reads were aligned to the reference genome (Bter_1.0, Refseq accession no. GCF_000214255.1 (Sadd et al., 2015)) using STAR v2.5.2 (Dobin et al., 2016) with standard parameters. Reads were counted per gene using HTseq v0.10.0 (Anders et al., 2015). Differential expression analysis was then carried out after count normalisation via a generalised linear model implemented by DEseq2 (Love et al., 2014) in R v.3.4.0 (http://www.R-project.org) with colony and reproductive status as variables. Genes were classed as differentially expressed when $q<0.05$ after correction for multiple testing using the Benjamini-Hochberg method (Benjamini and Hochberg, 1995).

Differential exon expression was determined using the same method for differential gene expression, using the R package DEXseq (Anders et al., 2012).

Differential methylation

BS-seq libraries were aligned to the reference genome (Bter_1.0, Refseq accession no. GCF_000214255.1, (Sadd et al., 2015)) using Bismark v.0.16.1 (Krueger and Andrews, 2011) and bowtie2 v.2.2.6 (Langmead and Salzberg, 2012) with standard parameters. Bismark was also used to extract methylation calls and carry out deduplication. Annotation of the methylation calls with genomic features (from the *B.terrestis* annotation file, Refseq accession no. GCF_000214255.1) was carried out using custom R scripts implementing the sqldf library (Grothendieck, 2017),

Prior to differential methylation analysis coverage outliers (above the 99.9th percentile) were removed along with bases covered by less than 10 reads. The methylation status of each CpG was then determined, using the 'methylation status calling' (MSC) procedure, as described in Cheng and Zhu (2014). CpG sites were then filtered to remove any site that did not return as methylated in at least one sample. This reduces the number of tests and hence decreases the stringency of the later FDR correction. This is a vital step for species with extremely low genome methylation where the

bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

majority of sites show zero methylation in all samples. A logistic regression model was then applied, via the R package methylKit (Akalin et al., 2012), to determine deferentially methylated sites, taking into account colony as a covariate. A minimum difference of at least 10% methylation and a q-value of <0.05 were required for a site to be classed as differentially methylated. We chose not to run a permutation as part of the differential methylation analysis, as seen in previous research (Arsenault et al., 2018; Libbrecht et al., 2016), as large variation found between colonies confounds the results, see supplementary 2.0.

GO analysis

Blast2GO v4.0.7 (Götz et al., 2008) was used to obtain associated GO terms for the *Drosophila melanogaster* homologs of differentially methylated genes and differentially expressed genes using blastx, allowing maximum target sequences of one with a minimum e-value of 1×10^{-3} . A custom R script was then used to carry out Fisher's exact test to identify significantly enriched GO terms, q<0.01 after Benjamini-Hochberg (Benjamini and Hochberg, 1995) correction for multiple testing. Treemaps were produced to visualise the GO terms using R scripts generated by REVIGO (Supek et al., 2011).

Comparative analyses

The hypergeometric test was applied to gene and GO term lists from the various analyses to determine if any overlap was statistically significant. Custom R scripts were used to investigate the relationship between gene expression and methylation. A reciprocal blast between the honeybee (Amel_4.5, Refseq accession no. GCA_000002195.1) and bumblebee genome (Bter_1.0, Refseq accession no. GCA_000214255.1) was carried out using blast+ v2.5.0 (Camacho et al., 2009), where the fasta sequence for each gene of each species was blasted against a custom database containing the fasta sequence for every gene of the opposite species, allowing only one match per gene and a minimum e-value of 1×10^{-3}. The results were then filtered to ensure only matches that occurred in both

bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

directions and to only one gene were used. For example multiple honeybee genes matched the same bumblebee gene, therefore all of these matches were discarded. A custom script was then used to check for overlap between the differentially methylated genes identified here and differentially methylated genes identified in Lyko et al. (2010) between honeybee reproductive castes.

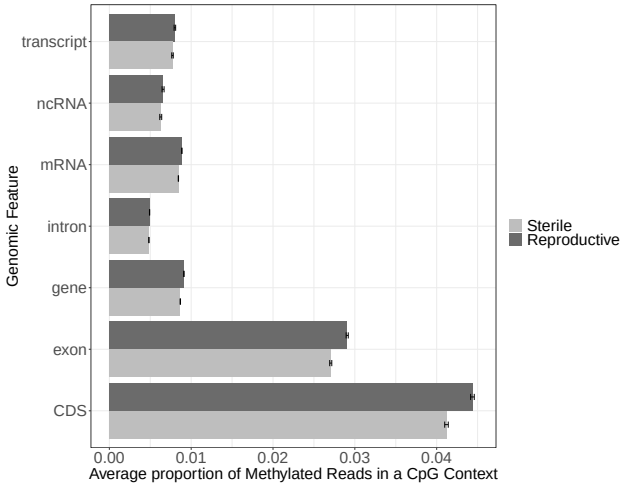
bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results

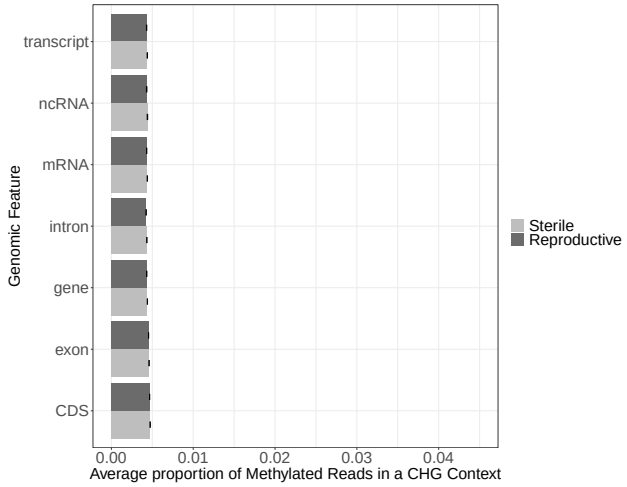
Genome-Wide Methylation Patterns

Up to a maximum of 10bp were trimmed from the start of all reads due to base bias generated by the Illumina sequencing protocol (Krueger et al., 2011). The mean mapping efficiency was 63.6% (+/-1.4%) and the mean coverage was 17.7 reads per base (+/-0.5). The percentage of methylated CpG's was a mean of 0.7% (+/-0.1%). Similar levels were also found in non-CpG contexts, a mean of 0.4% (+/-0.1%) of CHG's were methylated and 0.5% (+/-0.1%) CHH's, 'H' refers to any base other than guanine. The proportion of methylated reads for each C-context were then calculated (based on mean data from each replicate set) and the means, along with 95% confidence intervals, were plotted for various genomic features, based on the current *B.terrestris* annotation file (Refseq accession no. GCF_000214255.1). Introns were calculated based on exon and gene feature lengths and also included as an annotation. CpG methylation is enriched in coding regions and exons, and over four fold lower in introns (Fig.2a). CHG and CHH methylation however appear to be more evenly spread throughout the genome (Fig.2b and 2c).

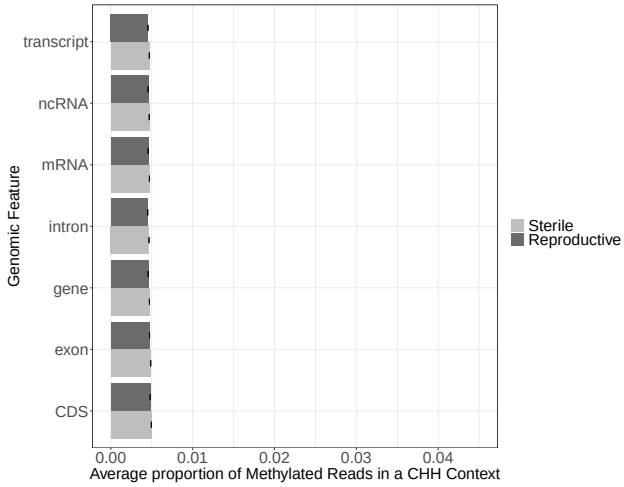
bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



(a) Genomic CpG Methylation



(b) Genomic CHG Methylation



(c) Genomic CHH Methylation

Figure 2: Proportion of methylated reads averaged from each reproductive status in three different genomic contexts: CpG (a), CHG (b) and CHH (c). 'H' referring to any base other then guanine. Error bars are 95% confidence intervals of the mean.

Methylation Differences Between Worker Reproductive Castes

A total of 624 CpGs found in 478 genes were identified as differentially methylated. However methylation differences were more apparent between colony replicates than between reproductive status, see supplementary 2.0. 202 of the genes identified as differentially methylated were hyper-

bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

177 methylated in reproductive workers and 225 were hyper-methylated in sterile workers. Additionally
178 51 genes contained CpGs hyper-methylated in reproductive workers as well as CpGs hyper-methylated
179 in sterile workers (supplementary 1.1). There were no differentially methylated sites found in a CHG
180 or CHH context.

181 Enriched GO terms for the *D.melanogaster* homologs of the 478 differentially methylated
182 genes include mostly cellular processes (supplementary 1.2), however there are also multiple GO
183 terms referring to behaviour (GO:0007610), reproductive processes (GO:0022414), reproduction
184 (GO:0000003), reproductive system development (GO:0061458) and gene expression (GO:0010467).

185 Similar GO terms are enriched when only the hyper-methylated genes of either reproductive
186 or sterile workers are assessed individually (supplementary 1.2). Interestingly however reproductive
187 processes (GO:0022414) and reproduction (GO:0000003) only occur enriched in the hyper-methylated
188 gene list for sterile workers.

189 Expression Differences Between Worker Reproductive Castes

190 All reads had 13bp trimmed from the start due to base bias generated by the Illumina protocol
191 (Krueger et al., 2011). The mean number of uniquely mapped reads was 89.4% (+/-0.8%). After
192 running a differential expression analysis with DESeq2, the decision was made to remove one sample
193 from all downstream analysis due to possible mislabelling of reproductive status (supplementary
194 2.1).

195 A total of 334 genes were identified as differentially expressed ($q<0.05$). 172 of these were
196 up-regulated in reproductive workers compared to sterile workers and 162 were up-regulated in
197 sterile workers compared to reproductive workers (supplementary 1.3).

198 One of the most up-regulated genes in reproductive workers was vitellogenin (gene ID:
199 LOC100650436, log2 fold-change of 2.92, $q = 4.85 \times 10^{-6}$). Previous work has found up-regulation
200 of this gene in reproductive *B.terrestris* workers is linked to aggressive behaviour rather than directly
201 to ovary development (Amsalem et al., 2014). Additionally two genes coding for serine-protease

bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

inhibitors were found to be up-regulated in reproductive workers, these proteins have been linked to reproduction in other insect species (Bao et al., 2014).

The GO term enrichment analysis of the differentially expressed genes shows enrichment for neuron differentiation (GO:0030182), imaginal disc morphogenesis (GO:0007560) and regulation of transcription (GO:1903506, GO:1903508, GO:0006357) amongst other cellular processes. Similar GO terms are also enriched when just the up-regulated genes of reproductive/sterile workers are assessed (supplementary 1.4).

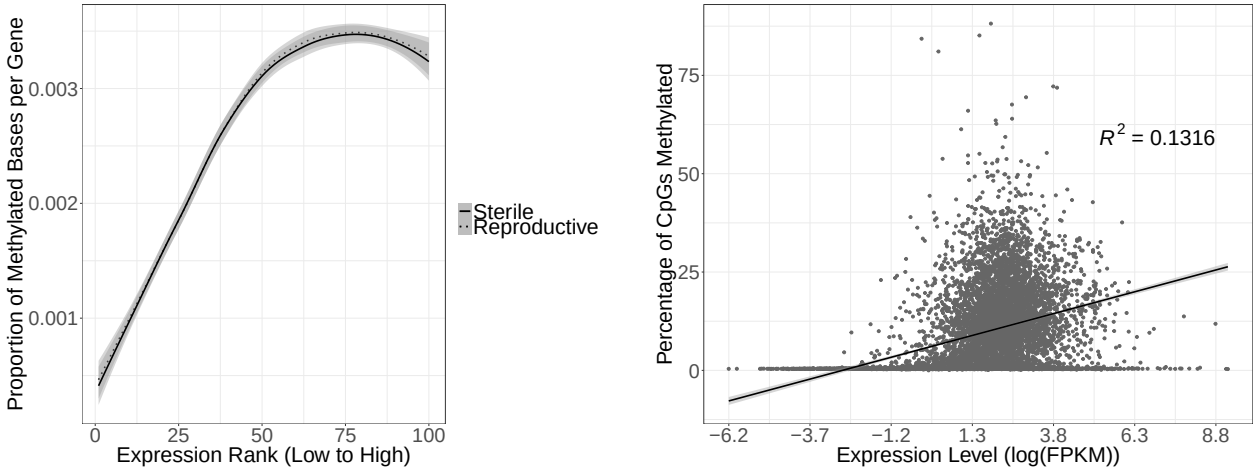
A total of 59 genes were identified as having differential exon usage, containing 83 differentially expressed exons ($q<0.1$, supplementary 1.5). From the 60 genes containing differentially expressed exons only 8 GO terms were enriched ($q<0.05$), the majority of which were involved in neuron processes (supplementary 1.6).

Relationship of Methylation and Gene Expression

The mean proportion of methylation per gene was plotted against ranked expression level ($\log_{10}$fpkm per gene) in 100 bins (from low to high) (Fig.3a) a linear model was fitted with reproductive status and expression level as independent variables. For all three groups (CpGs, CHHs, CHGs) there was no significant interaction between the effects of expression and reproductive status on methylation (interaction model versus main effects only model: CpG: $F_{1,197} = 0.001$, $p = 0.9747$, CHG: $F_{1,197} = 2.9675$, $p = 0.08653$, CHH: $F_{1,197} = 2.3113$, $p = 0.13$). There was a significant effect of expression on methylation in all contexts (CpG: $F_{1,197} = 727.8712$, $p = <2 \times 10^{-16}$, CHG: $F_{1,197} = 403.55$, $p = <2.2 \times 10^{-16}$, CHH: $F_{1,197} = 408.32$, $p = <2 \times 10^{-16}$). Reproductive and sterile castes were found to have similar levels of CpG methylation ($F_{1,197} = 0.0758$, $p = 0.7833$). However sterile workers had higher levels of CHG and CHH methylation compared to reproductive workers (CHG: $F_{1,197} = 32.763$, $p = 3.835 \times 10^{-8}$, CHH: $F_{1,197} = 35.415$, $p = 1.2 \times 10^{-8}$), supplementary 2.2.

There is also a positive correlation between expression and methylation level when the expression level (log FPKM) is plotted against the percentage of CpG's methylated per gene (Fig.3b).

bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



(a) (b)

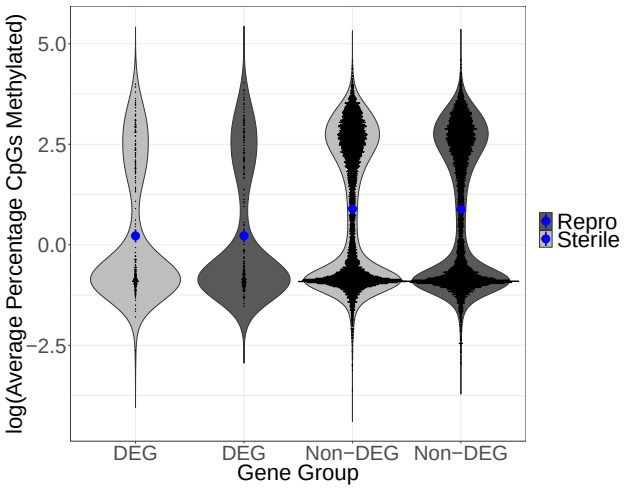
Figure 3: (a) Binned genes based on expression level (100 being the most highly expressed) plotted against the proportion of methylated bases in a CpG context. Data smoothed using the LOESS method, grey areas are 95% confidence intervals. (b) Scatter plot for CpG methylation level plotted against gene expression level (log-transformed fragments per kilobase per million), fitted with a linear regression. Each point represents an individual gene.

CpG methylation differences between differentially expressed genes and non-differentially expressed genes were assessed along with CpG methylation differences between genes containing differentially expressed exons and genes without differentially expressed exons (Fig.4a and Fig.4b). A linear model was fitted, as above, with reproductive status and gene context as independent variables. There was no significant interaction between gene context and reproductive status for either model (interaction versus main effects only model: differentially expressed genes: $F_{1, 22039} = 5 \times 10^{-4}$, p = 0.9818, differentially expressed exons: $F_{1, 343420} = 0.0092$, p = 0.9235). There was a significant effect on methylation depending on whether a gene was expressed or not and on whether a gene contained differentially expressed exons or not (differentially expressed genes: $F_{1, 22039} = 45.7942$, $p = 1.346 \times 10^{-11}$, differentially expressed exons: $F_{1, 343420} = 5.6118$, p = 0.01784). Methylation levels of genes in each context was not affected by reproductive status in either case (differentially expressed genes: $F_{1, 22039} = 0.0206$, p = 0.8859, differentially expressed exons: $F_{1, 343420} = 1.0646$, p = 0.30217).

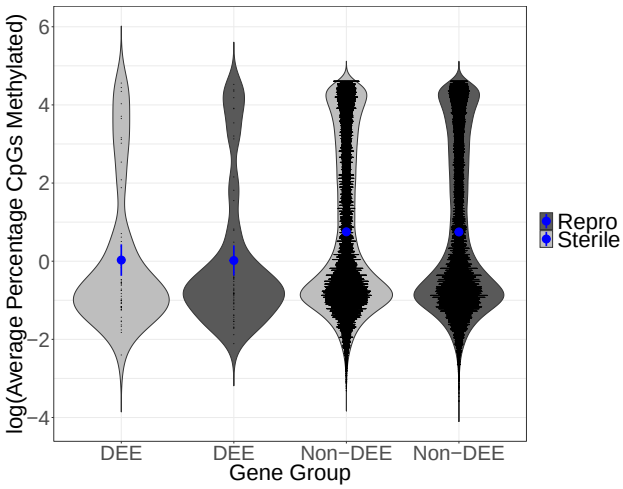
bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

CpG methylation levels of included or skipped exons was also examined (Fig.4c) using a linear model with reproductive status and exon status as independent variables. As above, there was no significant interaction between reproductive status and exon status ($F_{1, 139} = 0.0427$, $p = 0.8365$). There was also no significant effect of reproductive status or exon status on exon CpG methylation levels (reproductive status: $F_{1, 139} = 0.0278$, $p = 0.8677$, exon status: $F_{1, 139} = 0.0210$, $p = 0.8849$).

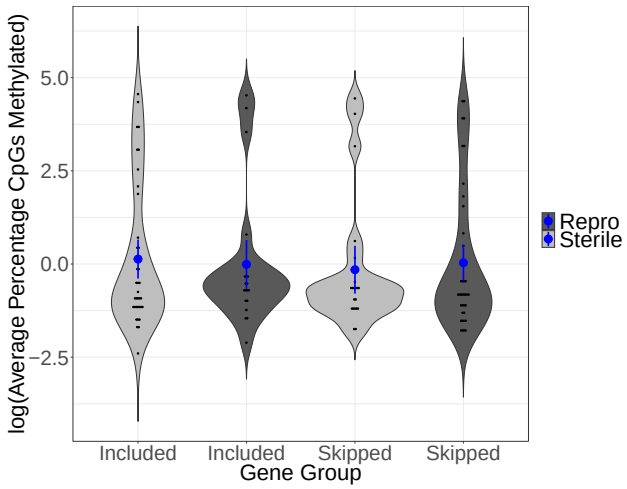
bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



(a) DEG .vs. Non-DEG



(b) DEE .vs. Non-DEE



(c) Included Exons .vs. Skipped Exons

Figure 4: Log transformed CpG methylation percentages of different sets of genes, represented with violin plots. These show the distribution of the data via a mirrored density plot, meaning the widest part of the plots represent the most genes. The blue dots represent the mean of each gene set along with error bars representing 95% confidence intervals of the mean. Each black dot is an individual gene or exon. DEG refers to differentially expressed genes and DEE refers to differentially expressed exons. Percentage methylation for figures (a) and (b) are over the entire gene and over the entire exon for figure (c).

Gene lists were checked for potential overlap from all analyses. There was no significant overlap between differentially methylated genes and differentially expressed genes (six genes, hypergeometric test p = 0.95). There was also no significant overlap between differentially methylated genes and

bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

248 genes containing differentially expressed exons (four genes, hypergeometric test p = 0.11).

249 There was a significant overlap of genes found to be differentially expressed and those
250 containing differentially expressed exons, 51 total (hypergeometric test, $p = 1.63 \times 10^{-64}$). All lists of
251 overlapping genes can be found in supplementary 1.7.

252 There was a significant overlap in GO terms between differentially methylated genes and
253 differentially expressed genes (hypergeometric test, $p = 2.29 \times 10^{-9}$, nine total). There was also a
254 significant overlap of GO terms between differentially methylated gene and genes with differentially
255 expressed exons (hypergeometric test, $p = 2.61 \times 10^{-5}$, two total) and between differentially expressed
256 gene and genes with differentially expressed exons (hypergeometric test, $p = 7.45 \times 10^{-27}$, eight total).
257 All lists of overlapping GO terms can be found in supplementary 1.8.

258 Honeybee Comparison

259 Custom honeybee and bumblebee homolog databases were created from 15,314 and 10,339 annotated
260 genes respectively (Amel_4.5 GCA_000002195.1, Bter_1.0 GCA_000214255.1). 9244 honeybee
261 genes matched at least one bumblebee gene and 7985 bumblebee genes matched at least one honeybee
262 gene with an e-value of $<1 \times 10^{-3}$. A total of 7345 genes made the same match in both blast searches.
263 Of these genes 392 matched more than one gene in one or both blasts and were therefore removed.
264 This left a final confident homolog list of 6953 genes. 417 of the 478 differentially methylated genes
265 identified here were present in the final homolog list however none of them matched the 549 genes
266 identified as differentially methylated between honeybee reproductive castes by Lyko et al. (2010).

bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Discussion

Using whole genome bisulfite sequencing and gene expression libraries from the same individual reproductive and sterile bumblebee workers, we have shown methylation is enriched in the coding regions of genes and that methylation has a weak but positive correlation with gene expression, as seen in other social insects (Patalano et al., 2015; Glastad et al., 2016; Bonasio et al., 2012). We also found differentially expressed genes contain lower levels of methylation compared to non-differentially expressed genes. We did not find a link between gene methylation and alternative splicing. Differentially methylated genes, found between the reproductive castes, contained GO terms enriched for reproductive mechanisms, among other biological processes. There was no significant overlap between genes that were differentially methylated and differentially expressed, although overlap between enriched GO terms for these gene sets indicates a potential interaction.

This is the first data set to accurately quantify methylation at base-pair resolution for *B.terrestris*. It confirms low methylation levels throughout the genome as predicted by Sadd et al. (2015). These low levels along with the enrichment for CpG methylation in coding regions is also seen in many social insect species, including *A.mellifera* (Lyko et al., 2010) and multiple ant species (Bonasio et al., 2012; Libbrecht et al., 2016). However this trend is not completely conserved among all social insects, for example the primitively social wasp species *Polistes dominula* shows 6% CpG methylation (Weiner et al., 2013), and the highly social termite, *Zootermopsis nevadensis*, has exceptionally high methylation levels, 12% CpG methylation, with methylation being just as common in introns as exons (Glastad et al., 2016).

Non-CpG methylation shows similar overall percentage levels in *B.terrestris* however it is not enriched in any particular genomic region. This suggests it may not have a regulatory role. Other social insect species have been found to have low levels of non-CpG methylation, again with no apparent enrichment in any particular genomic feature (Bonasio et al., 2012; Lyko et al., 2010). It is thought, in mammals, non-CpG methylation may be a bi-product of DNMT3a non-specific activity

bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(Ramsahoye et al., 2000). If a similar process were occurring in *B.terrestris*, this would explain the uniform low non-CpG methylation found across all genomic features. However significantly higher levels of non-CpG methylation was found to occur in sterile workers compared to reproductive workers (Fig.3a). Previous work which exposed *B.terrestris* to a chemical which should decrease general methylation levels also found workers were more likely to become reproductive (Amarasinghe et al., 2014). This indicates non-CpG methylation, whilst in lower general quantities, may still play a role in worker caste determination, although no differentially methylated non-CpG sites were found in this study.

Higher levels of non-CpG and CpG methylation are associated with higher levels of gene expression in *B.terrestris*. This is also the case in other social insects, with Figures 3a and 3b showing almost identical trends to those found in Bonasio et al. (2012), Patalano et al. (2015) and Libbrecht et al. (2016). Additionally other social insect species, show higher methylation in non-differentially expressed genes as we found here, examples include: *Dinoponera quadriceps* (Patalano et al., 2015), *Polistes canadensis* (Patalano et al., 2015), *Zootermopsis nevadensis* (Glastad et al., 2016) and *Cerapachys biroi* (Libbrecht et al., 2016). The higher levels of methylation in more highly expressed genes and in non-differentially expressed genes is thought to indicate a role for methylation in housekeeping genes in social insects (Foret et al., 2009; Lyko et al., 2010; Bonasio et al., 2012; Wang et al., 2013).

We found little evidence here for the role of exon methylation in alternative splicing. Previous research using ants and honeybees however did find an association between methylation and alternative splicing. A knock-down of DNMT3 by RNA interference was found to affect alternative splicing patterns in *A.mellifera*, with decreased methylation levels being directly related to exon skipping and intron retention (Li-Byarlay et al., 2013). Lower exon methylation was associated with exon skipping in two ant species; *Camponotus floridanus* and *Harpegnathos saltator* (Bonasio et al., 2012) and lower methylation was found in skipped and mutual exons, with higher methylation in retained introns in *C.biori* (Libbrecht et al., 2016). However other social insect species have also

bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

shown no direct association between methylation and alternative splicing (Wang et al., 2013).

Exon methylation has been shown to play role in histone modifications and nucleosome stability in mammals (Singer et al., 2015; Jones, 2012). These modifications have the ability to affect alternative splicing patterns through RNA polymerase accessibility, meaning whilst changes in DNA methylation may not be observed as directly related to alternative splicing, it's possible these changes have a downstream effect leading to transcriptional changes (Hunt et al., 2013). The analysis of the relationship between methylation and alternative splicing done here could be elaborated on further to include splicing sites and to also potentially identify the role of exon methylation in other epigenetic processes, which may themselves, lead to alternative splicing.

478 genes were found to have differential CpG methylation between reproductive and sterile worker castes. A worker classed as reproductive appeared to show a sterile transcriptional profile however and this was included in the pool for the reproductive sample for colony J8. This will have 'diluted' the strength of the methylation profile for this particular sample. It is therefore likely our data contains false negatives, meaning there may be differentially methylated genes between the reproductive castes which do not appear in our dataset. This may also have contributed to the lack of conserved differentially methylated genes identified between *B.terrestris* and *A.mellifera* reproductive castes. However the differentially methylated gene list obtained for *A.mellifera* used queen samples to represent the reproductive caste (Lyko et al., 2010), whereas here reproductive worker samples were used, this could also explain the lack of agreement. A further possibility is that each species has evolved methylation function independently, which is supported by the lack of phylogenetic concordance previously seen in insects (Bewick et al., 2016).

Gene ontology terms from the differentially methylated genes were enriched for behaviour and reproductive processes with the GO terms for reproduction occurring for hyper-methylated genes in sterile workers. As mentioned above lower levels of methylation have been associated with reproductive behaviour in *B.terrestis* (Amarasinghe et al., 2014). These results indicate methylation does play a mechanistic role in worker reproductive caste determination. However it is worth noting

bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

similar results have been found in other social insect species which also have additional epigenetic processes, such as microRNAs (Ashby et al., 2016) contributing independently to caste determination. Additionally Simola et al. (2016) found histone acetylation differences between worker castes of *Camponotus floridanus*, they inhibited histone acetylation and found this caused the major worker caste to behave more like a minor worker. This same species has also been shown to have caste specific methylation profiles (Bonasio et al., 2012).

These examples indicate it is likely an interplay between multiple mechanisms that ultimately cause social insect caste differentiation. This idea is also supported by the fact that here, we find very little overlap between differentially methylated genes and differentially expressed genes, however their corresponding enriched GO terms do overlap. The majority of these terms are related to neuron function. Interestingly non-CpG methylation in mammals has been associated with gene expression changes in neurons (Tognini et al., 2015). Whilst non-CpG methylation in *B.terrestris* appeared to be randomly spread throughout the genome, there was still a significant difference between levels of sterile and reproductive workers. Future work should attempt to identify where changes in non-CpG methylation occur within the genome and if there is any relationship between these sites and gene expression or other epigenetic modifications.

Overall the *B.terrestris* methylome appears similar to most other social insects, in terms of overall levels and association with gene expression. We found many differentially methylated genes between reproductive castes, which combined with previous research, (Amarasinghe et al., 2014), indicates a role for methylation in reproductive caste determination in *B.terrestris*. However it is clear, owing to the lack of consistency between differentially methylated genes and differentially expressed genes, methylation is not directly responsible for the associated changes in gene expression leading to the different reproductive phenotypes in *B.terrestris*. It is possible methylation is just one of multiple mechanism working in unison to define the different caste phenotypes in *B.terrestris*. Future work should focus on the consequences of experimental methylation removal or addition (Pegoraro et al., 2017), as well as exploring additional epigenetic mechanisms to attempt to identify

bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

a full pathway leading to reproductive caste differences.

This study provides new insights into the nature of a bumblebee methylome and it's interaction with gene expression and alternative splicing. We also found methylation contributes to reproductive caste plasticity in *B.terrestris*. These findings contribute to our understanding of the role of methylation in social insects.

Acknowledgements

Thank you to Dr. Ben Hunt for important advice and discussion regarding data analysis. Thank you to Dr. Ezio Rosato for providing valuable discussion. This research used the ALICE2 High Performance Computing Facility at the University of Leicester. HM was supported by a NERC CENTA DTP studentship. ZL was supported by a BBSRC MIBTP DTP studentship. EBM was funded by NERC grant NE/N010019/1.

Author Contributions

E.B.M. conceived the study. H.M. and Z.N.L. conducted the experiment. H. M. analysed the data and wrote the initial manuscript. All authors contributed to and reviewed the final manuscript.

Data Accessibility

Data will be deposited in GenBank upon paper acceptance. All code will be also be made available at: DOI: 10.5281/zenodo.2394171.

bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

Akalin, A., M. Kormaksson, S. Li, F. E. Garrett-bakelman, M. E. Figueroa, A. Melnick, and C. E. Mason
2012. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome biology*, 13(R87).

Amarasinghe, H. E., C. I. Clayton, and E. B. Mallon
2014. Methylation and worker reproduction in the bumble-bee (Bombus terrestris). *Proceedings of the Royal Society B: Biological Sciences*, 281(20132502).

Amsalem, E., O. Malka, C. Grozinger, and A. Hefetz
2014. Exploring the role of juvenile hormone and vitellogenin in reproduction and social behavior in bumble bees. *BMC Evolutionary Biology*, 14(1):45.

Anders, S., P. T. Pyl, and W. Huber
2015. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2):166–169.

Anders, S., A. Reyes, and W. Huber
2012. Detecting differential usage of exons from RNA-seq data. *Genome Research*, 22(10):2008–2017.

Arsenault, S. V., B. G. Hunt, and S. M. Rehan
2018. The effect of maternal care on gene expression and DNA methylation in a subsocial bee. *Nature Communications*, 9(3468).

Ashby, R., S. Forêt, I. Searle, and R. Maleszka
2016. MicroRNAs in Honey Bee Caste Determination. *Scientific Reports*, 6(18794).

Bao, Y. Y., X. Qin, B. Yu, L. B. Chen, Z. C. Wang, and C. X. Zhang
2014. Genomic insights into the serine protease gene family and expression profile analysis in the planthopper, Nilaparvata lugens. *BMC Genomics*, 15(1).

Benjamini, Y. and Y. Hochberg
1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, 57(1):289–300.

Bewick, A. J., K. J. Vogel, A. J. Moore, and R. J. Schmitz
2016. Evolution of DNA methylation across insects. *Molecular Biology and Evolution*, 34(3):654–665.

Bonasio, R., Q. Li, J. Lian, N. S. Mutti, L. Jin, H. Zhao, P. Zhang, P. Wen, H. Xiang, Y. Ding, Z. Jin, S. S. Shen, Z. Wang, W. Wang, J. Wang, S. L. Berger, J. J. Liebig, G. Zhang, and D. Reinberg
2012. Genome-wide and caste-specific DNA methylomes of the ants camponotus floridanus and harpegnathos saltator. *Current Biology*, 22(19):1755–1764.

bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden
2009. BLAST+: Architecture and applications. *BMC Bioinformatics*, 10:1–9.

Cheng, L. and Y. Zhu
2014. A classification approach for DNA methylation profiling with bisulfite next-generation sequencing data. *Bioinformatics*, 30(2):172–179.

Chevin, L.-M., R. Lande, and G. M. Mace
2010. Adaptation, Plasticity, and Extinction in a Changing Environment: Towards a Predictive Theory. *PLoS Biology*, 8(4):e1000357.

Collins, D. H., I. Mohorianu, M. Beckers, V. Moulton, T. Dalmay, and A. F. Bourke
2017. MicroRNAs Associated with Caste Determination and Differentiation in a Primitively Eusocial Insect. *Scientific Reports*, 7(45674).

Dobin, A., T. R. Gingeras, and C. Spring
2016. Mapping RNA-seq Reads with STAR Alexander. *Current Protocols in Bioinformatics*, (51):1–11.

Duchateau, M. J. and H. H. W. Velthuis
1988. Development and reproductive strategies in Bombus terrestris colonies. *Behaviour*, 107(3):186–207.

Fang, F., E. Hodges, a. Molaro, M. Dean, G. J. Hannon, and a. D. Smith
2012. Genomic landscape of human allele-specific DNA methylation. *Proceedings of the National Academy of Sciences*, 109(19):7332–7337.

Feng, S., S. J. Cokus, X. Zhang, P.-Y. Chen, M. Bostick, M. G. Goll, J. Hetzel, J. Jain, S. H. Strauss, M. E. Halpern, C. Ukomadu, K. C. Sadler, S. Pradhan, M. Pellegrini, and S. E. Jacobsen
2010. Conservation and divergence of methylation patterning in plants and animals. *Proceedings of the National Academy of Sciences*, 107(19):8689–8694.

Foret, S., R. Kucharski, Y. Pittelkow, G. A. Lockett, and R. Maleszka
2009. Epigenetic regulation of the honey bee transcriptome: unravelling the nature of methylated genes. *BMC Genomics*, 10(1):472.

Foster, R. L., A. Brunskill, D. Verdirame, and S. O'Donnell
2004. Reproductive physiology, dominance interactions, and division of labour among bumble bee workers. *Physiological Entomology*, 29(4):327–334.

Geva, S., K. Hartfelder, and G. Bloch
2005. Reproductive division of labor, dominance, and ecdysteroid levels in hemolymph and ovary of the bumble bee Bombus terrestris. *Journal of Insect Physiology*, 51(7):811–823.

Glastad, K. M., L. M. Chau, and M. A. Goodisman
2015. Epigenetics in Social Insects. In *Physiology, Behavior, Genomics of Social Insects*, volume 48, Pp. 227–269. Elsevier Ltd.

bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

458 Glastad, K. M., K. Gokhale, J. Liebig, and M. A. D. Goodisman
459 2016. The caste- and sex-specific DNA methylome of the termite Zootermopsis nevadensis.
460 *Scientific Reports*, 6(37110).

461 Götz, S., J. M. García-Gómez, J. Terol, T. D. Williams, S. H. Nagaraj, M. J. Nueda, M. Robles,
462 M. Talón, J. Dopazo, and A. Conesa
463 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic*
464 *Acids Research*, 36(10):3420–3435.

465 Grothendieck, G.
466 2017. sqldf: Manipulate R Data Frames Using SQL. R package:https://cran.r–
467 project.org/package=sqldf.

468 Herb, B. R., F. Wolschin, K. D. Hansen, M. J. Aryee, B. Langmead, R. Irizarry, G. V. Amdam, and
469 A. P. Feinberg
470 2012. Reversible switching between epigenetic states in honeybee behavioral subcastes. *Nature*
471 *Neuroscience*, 15(10):1371–1373.

472 Hunt, B. G., K. M. Glastad, S. V. Yi, and M. A. D. Goodisman
473 2013. The function of intragenic DNA methylation: Insights from insect epigenomes. *Integrative*
474 *and Comparative Biology*, 53(2):319–328.

475 Jones, P. A.
476 2012. Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nature Reviews*
477 *Genetics*, 13(7):484–492.

478 Krueger, F. and S. R. Andrews
479 2011. Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinfor-*
480 *matics*, 27(11):1571–1572.

481 Krueger, F., S. R. Andrews, and C. S. Osborne
482 2011. Large scale loss of data in low-diversity illumina sequencing libraries can be recovered by
483 deferred cluster calling. *PLoS ONE*, 6(1):4–10.

484 Kucharski, R., J. Maleszka, S. Foret, and R. Maleszka
485 2008. Nutritional control of reproductive status in honeybees via DNA methylation. *Science*,
486 319(5871):1827–1830.

487 Langmead, B. and S. L. Salzberg
488 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods*, 9(4):357–359.

489 Li, B., L. Hou, D. Zhu, X. Xu, S. An, and X. Wang
490 2018. Identification and caste-dependent expression patterns of DNA methylation associated
491 genes in Bombus terrestris. *Scientific Reports*, 8(2332).

bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Li-Byarlay, H., Y. Li, H. Stroud, S. Feng, T. C. Newman, M. Kaneda, K. K. Hou, K. C. Worley, C. G. Elsik, S. A. Wickline, S. E. Jacobsen, J. Ma, and G. E. Robinson
2013. RNA interference knockdown of DNA methyl-transferase 3 affects gene alternative splicing in the honey bee. *Proceedings of the National Academy of Sciences*, 110(31):12750–12755.

Libbrecht, R., P. R. Oxley, L. Keller, and D. J. C. Kronauer
2016. Robust DNA Methylation in the Clonal Raider Ant Brain. *Current Biology*, 26:1–5.

Lonsdale, Z., K. Lee, M. Kiriakidu, H. Amarasinghe, D. Nathanael, C. J. O'Connor, and E. B. Mallon
2017. Allele specific expression and methylation in the bumblebee, Bombus terrestris. *PeerJ*, 5(e3798).

Love, M. I., W. Huber, and S. Anders
2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, 15(12):550.

Lyko, F., S. Foret, R. Kucharski, S. Wolf, C. Falckenhayn, and R. Maleszka
2010. The honey bee epigenomes: Differential methylation of brain DNA in queens and workers. *PLoS Biology*, 8(11).

Martin, M.
2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):10.

Matsuura, K., N. Mizumoto, K. Kobayashi, T. Nozaki, T. Fujita, T. Yashiro, T. Fuchikawa, Y. Mitaka, and E. L. Vargo
2018. A Genomic Imprinting Model of Termite Caste Determination: Not Genetic but Epigenetic Inheritance Influences Offspring Caste Fate. *The American Naturalist*, 191(6).

Mott, B. M., J. Gadau, and K. E. Anderson
2015. Phylogeography of Pogonomyrmex barbatus and P. rugosus harvester ants with genetic and environmental caste determination. *Ecology and Evolution*, 5(14):2798–2826.

Patalano, S., A. Vlasova, C. Wyatt, P. Ewels, F. Camara, P. G. Ferreira, C. L. Asher, T. P. Jurkowski, A. Segonds-Pichon, M. Bachman, I. González-Navarrete, A. E. Minoche, F. Krueger, E. Lowy, M. Marcet-Houben, J. L. Rodriguez-Ales, F. S. Nascimento, S. Balasubramanian, T. Gabaldon, J. E. Tarver, S. Andrews, H. Himmelbauer, W. O. H. Hughes, R. Guigó, W. Reik, and S. Sumner
2015. Molecular signatures of plastic phenotypes in two eusocial insect species with simple societies. *Proceedings of the National Academy of Sciences*, 112(45):13970–13975.

Pegoraro, M., H. Marshall, Z. N. Lonsdale, and E. B. Mallon
2017. Do social insects support Haig's kin theory for the evolution of genomic imprinting? *Epigenetics*, 12(9):725–742.

bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Ramsahoye, B. H., D. Biniszkiewicz, F. Lyko, V. Clark, a. P. Bird, and R. Jaenisch
2000. Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proceedings of the National Academy of Sciences of the United States of America*, 97(10):5237–5242.

Sadd, B. M., S. M. Barribeau, G. Bloch, D. C. de Graaf, P. Dearden, C. G. Elsik, J. Gadau, C. J. Grimmelikhuijzen, M. Hasselmann, J. D. Lozier, H. M. Robertson, G. Smagghe, E. Stolle, M. Van Vaerenbergh, R. M. Waterhouse, E. Bornberg-Bauer, S. Klasberg, A. K. Bennett, F. Câmara, R. Guigó, K. Hoff, M. Mariotti, M. Munoz-Torres, T. Murphy, D. Santesmasses, G. V. Amdam, M. Beckers, M. Beye, M. Biewer, M. M. Bitondi, M. L. Blaxter, A. F. Bourke, M. J. Brown, S. D. Buechel, R. Cameron, K. Cappelle, J. C. Carolan, O. Christiaens, K. L. Ciborowski, D. F. Clarke, T. J. Colgan, D. H. Collins, A. G. Cridge, T. Dalmay, S. Dreier, L. du Plessis, E. Duncan, S. Erler, J. Evans, T. Falcon, K. Flores, F. C. Freitas, T. Fuchikawa, T. Gempe, K. Hartfelder, F. Hauser, S. Helbing, F. C. Humann, F. Irvine, L. S. Jermiin, C. E. Johnson, R. M. Johnson, A. K. Jones, T. Kadowaki, J. H. Kidner, V. Koch, A. Köhler, F. B. Kraus, H. M. G. Lattorff, M. Leask, G. A. Lockett, E. B. Mallon, D. S. Antonio, M. Marxer, I. Meeus, R. F. Moritz, A. Nair, K. Näpflin, I. Nissen, J. Niu, F. M. Nunes, J. G. Oakeshott, A. Osborne, M. Otte, D. G. Pinheiro, N. Rossié, O. Rueppell, C. G. Santos, R. Schmid-Hempel, B. D. Schmitt, C. Schulte, Z. L. Simões, M. P. Soares, L. Swevers, E. C. Winnebeck, F. Wolschin, N. Yu, E. M. Zdobnov, P. K. Aqrawi, K. P. Blankenburg, M. Coyle, L. Francisco, A. G. Hernandez, M. Holder, M. E. Hudson, L. R. Jackson, J. Jayaseelan, V. Joshi, C. Kovar, S. L. Lee, R. Mata, T. Mathew, I. F. Newsham, R. Ngo, G. Okwuonu, C. Pham, L. L. Pu, N. Saada, J. Santibanez, D. N. Simmons, R. Thornton, A. Venkat, K. K. Walden, Y. Q. Wu, G. Debyser, B. Devreese, C. Asher, J. Blommaert, A. D. Chipman, L. Chittka, B. Fouks, J. Liu, M. P. O'Neill, S. Sumner, D. Puiu, J. Qu, S. L. Salzberg, S. E. Scherer, D. M. Muzny, S. Richards, G. E. Robinson, R. A. Gibbs, P. Schmid-Hempel, and K. C. Worley
2015. The genomes of two key bumblebee species with primitive eusocial organization. *Genome Biology*, 16(1):76.

Schneider, C. A., W. S. Rasband, and K. W. Eliceiri
2012. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9(7):671–675.

Shaham, R., R. Ben-Shlomo, U. Motro, and T. Keasar
2016. Genome methylation patterns across castes and generations in a parasitoid wasp. *Ecology and Evolution*, 6(22):7943–7953.

Simola, D. F., R. J. Graham, C. M. Brady, B. L. Enzmann, C. Desplan, A. Ray, L. J. Zwiebel, R. Bonasio, D. Reinberg, J. Liebig, and S. L. Berger
2016. Epigenetic (re)programming of caste-specific behavior in the ant Camponotus floridanus. *Science*, 351(6268).

Singer, M., I. Kosti, L. Pachter, and Y. Mandel-Gutfreund
2015. A diverse epigenetic landscape at human exons with implication for expression. *Nucleic Acids Research*, 43(7):3498–3508.

bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Standage, D. S., A. J. Berens, K. M. Glastad, A. J. Severin, V. P. Brendel, and A. L. Toth
2016. Genome, transcriptome, and methylome sequencing of a primitively eusocial wasp reveal a greatly reduced DNA methylation system in a social insect. *Molecular Ecology*, 25(8):1769–1784.

Supek, F., M. Bošnjak, N. Škunca, and T. Šmuc
2011. Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS ONE*, 6(7).

Tognini, P., D. Napoli, and T. Pizzorusso
2015. Dynamic DNA methylation in the brain: a new epigenetic mark for experience-dependent plasticity. *Frontiers in Cellular Neuroscience*, 9(331).

Wang, X., D. Wheeler, A. Avery, A. Rago, J. H. Choi, J. K. Colbourne, A. G. Clark, and J. H. Werren
2013. Function and Evolution of DNA Methylation in Nasonia vitripennis. *PLoS Genetics*, 9(10).

Weiner, S. A., D. A. Galbraith, D. C. Adams, N. Valenzuela, F. B. Noll, C. M. Grozinger, and A. L. Toth
2013. A survey of DNA methylation across social insect species, life stages, and castes reveals abundant and caste-associated methylation in a primitively social wasp. *Naturwissenschaften*, 100(8):795–799.

Winkler, A. M., M. A. Webster, D. Vidaurre, T. E. Nichols, and S. M. Smith
2015. Multi-level block permutation. *NeuroImage*, 123:253–268.

Woodard, S. H., J. D. Lozier, D. Goulson, P. H. Williams, J. P. Strange, and S. Jha
2015. Molecular tools and bumble bees: Revealing hidden details of ecology and evolution in a model system. *Molecular Ecology*, 24(12):2916–2936.

bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary

2.0: Permutation Analysis

Permutation tests are used to randomly shuffle labels across a given data set to ensure results obtained via a significance test are due to biological causes rather than random variation within the data resulting in type I errors.

Here we employed a similar method to Arsenault et al. (2018) and Libbrecht et al. (2016). After removal of positions containing zero methylation for every sample and filtering by coverage, the sample labels for data at each CpG were randomly shuffled 10,000 times. A logistic regression was carried out for each random data set using methylKit. The number of differentially methylated sites per permutation were plotted (Fig.5).

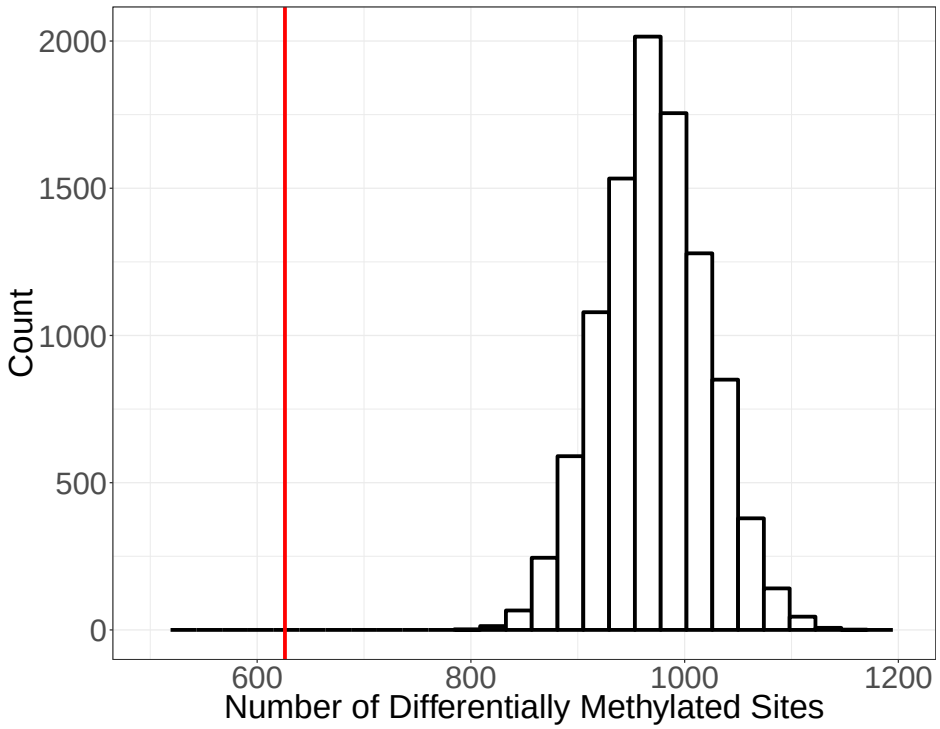
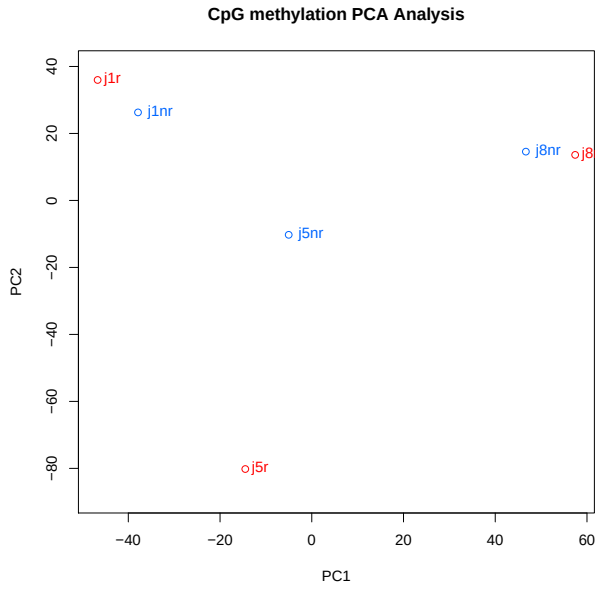


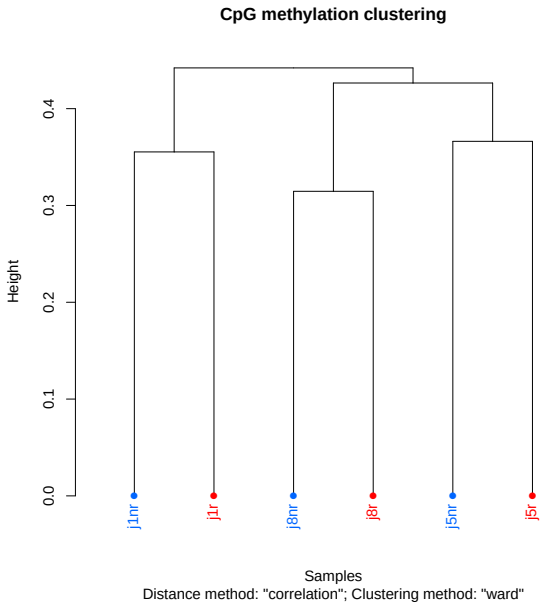
Figure 5: Histogram of the number of differentially methylated sites obtained from 10,000 permutations. The red line indicates the number of sites observed from the un-shuffled data.

More sites are found to be differentially methylated using the random data as variation between colonies is higher than variation caused by reproductive status (Fig.6a and 6b). Using this method 'colony' can no longer be taken into account as a consistent covariate and the effect therefore inflates the number of obtained differentially methylated sites.

bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



(a) PCA Plot



(b) Dendogram

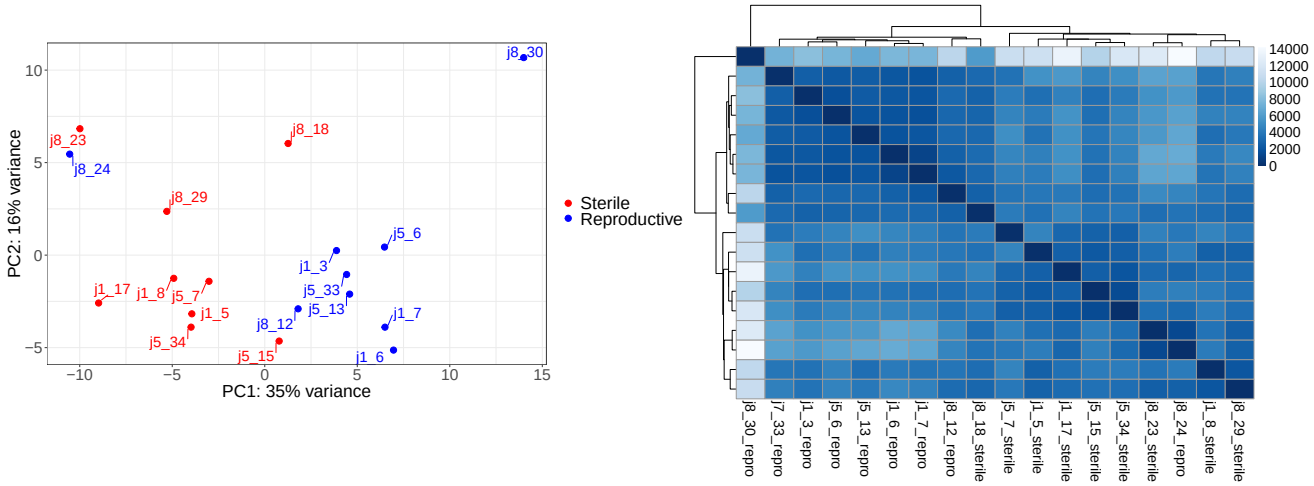
Figure 6: (a) PCA plot showing samples cluster more closely by colony than by reproductive status. (b) Dendogram showing sample cluster by colony. Red labels indicate reproductive samples and blue labels represent sterile samples.

Whilst permutation tests are useful for some data sets, when structure is present in the data they become unreliable, as discussed in Winkler et al. (2015). A higher number of replicates would allow label shuffling within confounding factors, maintaining the structure of the data, thus allowing a valid permutation.

2.1: Sample Exclusion from Differential Expression

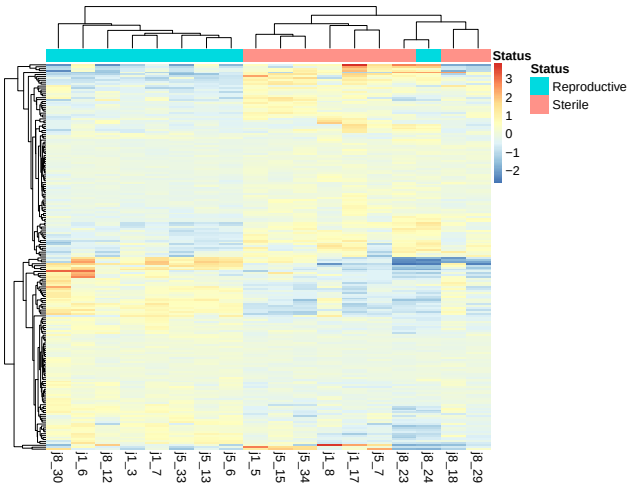
Sample J8_24 was classed as reproductive but clustered with the sterile samples in both a principle component analysis (PCA) and a poisson distance matrix, taking into account expression levels of all genes (Fig.7a and 7b). It also clustered with sterile samples in a hierarchical cluster using euclidean distances based on the top 100 differentially expressed genes (Fig.7c). Including this sample led to a decrease in the number of differentially expressed genes identified (110 with sample, 334 without sample). After removal of this sample all other samples clustered by reproductive status (Fig.8a, 8b and 8c).

bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



(a) PCA Plot

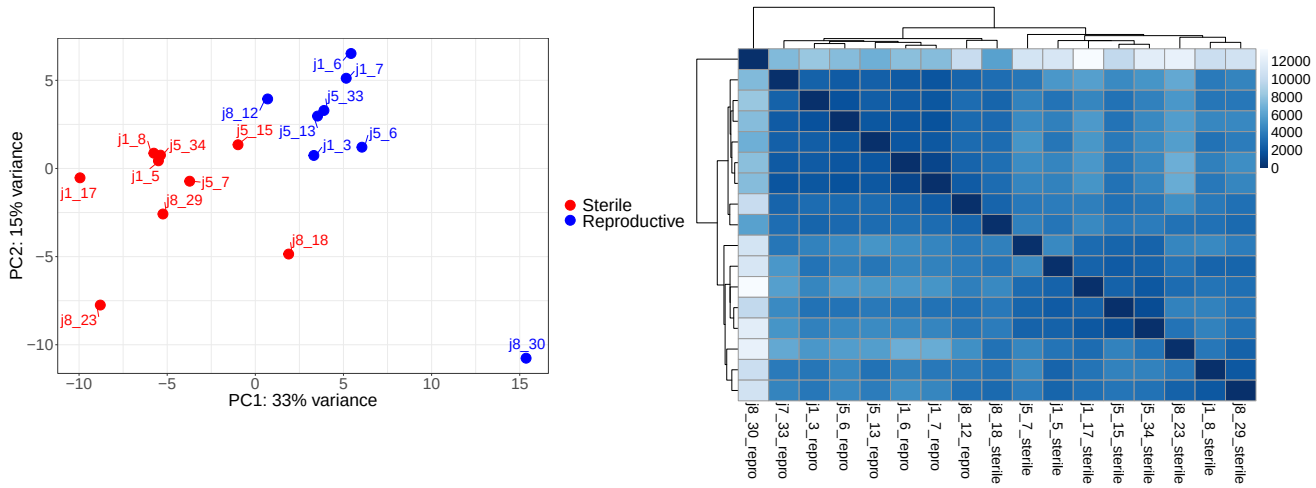
(b) Poisson Distance Matrix



(c) Elucidean Cluster Heatmap

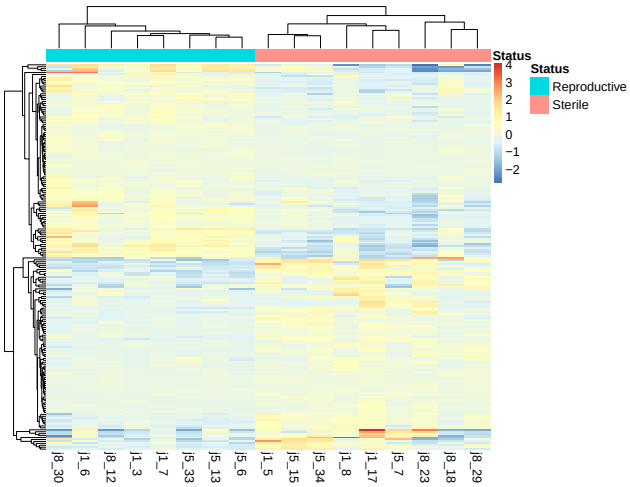
Figure 7: Graphs generated from differential expression analysis using DESEQ2 for all samples. (a) PCA plot showing the different reproductive status' grouped together. (b) Poisson measurement of dissimilarity between counts. (c) Heatmap showing the top 100 differentially expressed genes in reproductive and sterile workers. Red colour indicates over-expression and blue colour indicates under-expression.

bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



(a) PCA Plot

(b) Poisson Distance Matrix

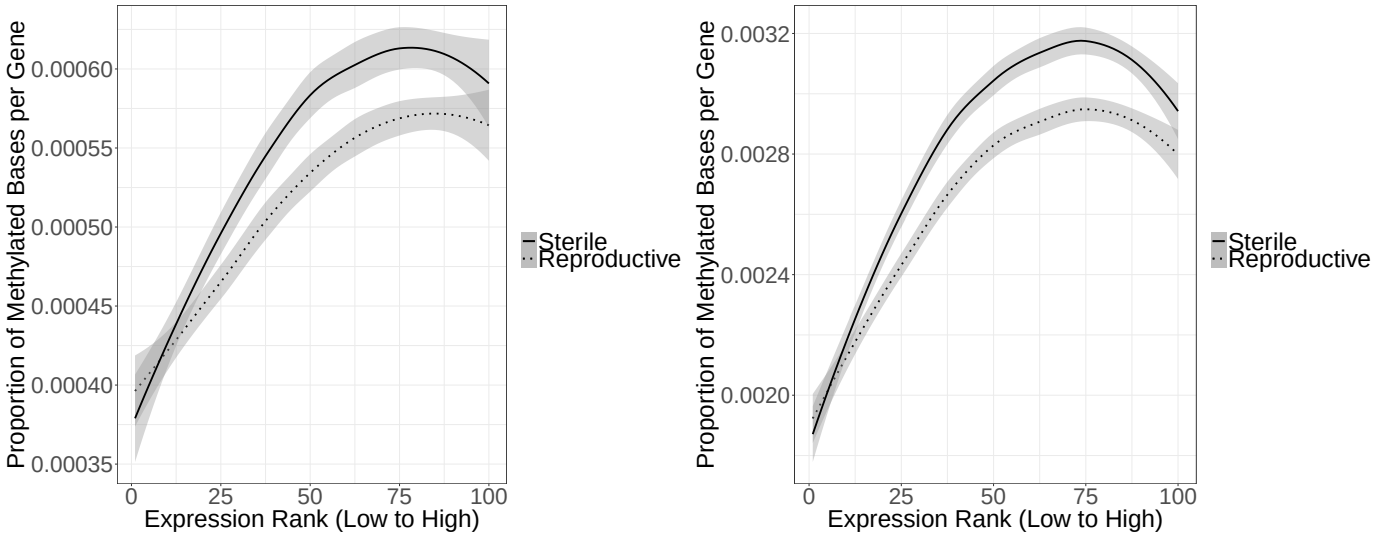


(c) Elucidean Cluster Heatmap

Figure 8: Graphs generated from differential expression analysis using DESEQ2 for all samples, excluding J8_24. (a) PCA plot showing the different reproductive status' grouped together. (b) Poisson measurement of dissimilarity between counts. (c) Heatmap showing the top 100 differentially expressed genes in reproductive and sterile workers. Red colour indicates over-expression and blue colour indicates under-expression.

bioRxiv preprint doi: https://doi.org/10.1101/517698; this version posted January 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2.2: Non-CpG Analysis Graphs



(a) CHG (b) CHH

Figure 9: Binned genes based on expression level (100 being the most highly expressed) plotted against the proportion of methylated bases in a (a) CHG and (b) CHH context. Data smoothed using the LOESS method, grey areas are 95% confidence intervals.