Genome-wide disruption of DNA methylation by 5-aza-2'-deoxycytidine in the parasitoid wasp *Nasonia vitripennis*

Cook N[1]*, Parker DJ[2,3]*, Turner F[4], Tauber E[5], Pannebakker BA[6] and Shuker DM[1]

*These authors contributed equally

1. School of Biology, University of St Andrews, UK

2. Department of Ecology and Evolution, University of Lausanne, Switzerland

3. Swiss Institute of Bioinformatics, Lausanne, Switzerland

4. Edinburgh Genomics, School of Biological Sciences, University of Edinburgh, UK

5. Faculty of Natural Sciences, University of Haifa, Israel

6. Laboratory of Genetics, Wageningen University, The Netherlands

Keywords: 5-aza-2'-deoxycytidine, 5-aza-dC, DNA methylation, epigenetics, *Nasonia vitripennis*, sex allocation

Abstract

DNA methylation of cytosine residues across the genome influences how genes and phenotypes are regulated in a wide range of organisms. As such, understanding the role of DNA methylation and other epigenetic mechanisms has become very much a part of mapping genotype to phenotype, a major question in evolutionary biology. Ideally, we would like to manipulate DNA methylation patterns on a genome-wide scale, to help us to elucidate the role that epigenetic modifications play in phenotypic expression. Recently, the demethylating agent 5-aza-2'-deoxycytidine (5-aza-dC; commonly used in the epigenetic treatment of certain cancers), has been deployed to explore the epigenetic regulation of a number of traits of interest to evolutionary ecologists, including facultative sex allocation in the parasitoid wasp *Nasonia vitripennis*. In a recent study, we showed that treatment with 5-aza-dC did not ablate the facultative sex allocation response in *Nasonia*, but shifted the patterns of sex allocation in a way predicted by genomic conflict theory. This was the first (albeit indirect) experimental evidence for genomic conflict over sex allocation facilitated by DNA methylation. However, that work lacked direct evidence of the effects of 5-aza-dC on DNA methylation, and indeed the effect of the chemical has since been questioned in *Nasonia*. Here, using whole-genome bisulphite sequencing of more than 4 million CpGs, across more than 11,000 genes, we demonstrate unequivocally that 5-aza-dC disrupts methylation on a large scale across the *Nasonia vitripennis* genome. We show that the disruption can lead to both hypo- and hyper-methylation, may vary across tissues and time of sampling, and that the effects of 5-aza-dC are context- and sequence specific. We conclude that 5-aza-dC does indeed have the potential to be repurposed as a tool for studying the role of DNA methylation in evolutionary ecology, whilst many details of its action remain to be discovered.

Introduction

The genomics revolution has transformed our understanding of how genomes build phenotypes. The extraordinary diversity of morphologies, behaviours, and physiological adaptations we see across the natural world are not necessarily the product of very different sets of genes, or very different numbers of genes in organisms, but rather differences in how, where and when genes are expressed during development. The genomics revolution has therefore been as much about understanding gene regulation as it has been about identifying allelic variants at genetic loci (Ashbrook *et al.* 2017, Biémont *et al.* 2010). At the forefront of our growing knowledge of gene regulation has been the role of epigenetic modification of DNA, including DNA methylation. In the vast majority of cases, DNA methylation occurs at the C5 position of the cytosine ring in CpG dinucleotides and is catalysed by DNA methyltransferase enzymes (Dnmts). Three types of Dnmt are known in eukaryotes; Dnmt3 establishes DNA methylation patterns *de novo*, Dnmt1 maintains these patterns, and Dnmt2 is involved in tRNA methylation (Lyko 2018). In mammals, broadly, DNA methylation is found throughout the genome (except at CpG islands near the promoters of genes) and is associated with transcriptional repression (Feng *et al.* 2010, Zemach *et al.* 2010). By contrast, in insects, DNA methylation is concentrated in gene bodies and is associated with more stable patterns of gene expression (Bewick *et al.* 2017). Notably, vertebrate and invertebrate DNA methylation varies greatly in its extent throughout the genome, with the proportion of methylated CpGs much lower in invertebrates; 0-14% as compared to 60-90% in mammals (Bewick *et al.* 2017, Glastad *et al.* 2011). The evolutionary reasons for these differences are not yet clear, although it is thought that gene body methylation is an ancient feature of eukaryote genomes (Zemach *et al.* 2010). However, DNA methylation status influences how genes (and phenotypes) are expressed in many organisms, including reproductive vs non-reproductive worker bumblebees (Amarasinghe *et al.* 2014), caste differentiation (Kucharski *et al.* 2008, Lyko *et al.* 2010), and learning and memory (Li *et al.* 2017) in honey bees, and behavioural division of labour in ants (*Pogonomyrmex barbatus*; Smith *et al.* 2012; but see Libbrecht et al. 2016 for a robust critique of earlier studies of DNA methylation and caste determination in social insects). Therefore, DNA methylation patterns are very much part of mapping genotypes to phenotypes.

The flexible control of gene expression provided by mechanisms such as DNA methylation has also caught the interest of biologists interested in the underlying mechanisms of phenotypic plasticity, and indeed for evolutionary biologists interested in how such plasticity may evolve (Bonduriansky and Day 2009, 2018). Moreover, there has also been great interest in the extent to which epigenetic modifications may be inherited across generations (Lind and Spagopoulou 2018, Burggren 2016, Bondurianksy and Day 2009). Transmission of epigenetic state allows transmission of parent-of-origin information, or so-called genomic imprinting, setting the scene for genomic conflict (Patten *et al.*

74 2014, Haig 2002). Genomic conflict arises when maternally- and paternally- inherited alleles have
75 different evolutionary optima for traits expressed in an individual (Haig 1997, 2002). A classic example
76 is the expression of maternally- or paternally- inherited alleles during embryonic development in
77 mammals: paternally-inherited alleles are selected to extract more resources from the mother *in utero*
78 compared to maternally-inherited alleles (Moore and Haig 1991). This conflict is now known to be, at
79 least in part, facilitated by parent-of-origin specific gene expression of insulin-like growth factor II
80 (Igf2) and Igf2r. In mammals, the Igf2 gene promotes growth and cellular differentiation during
81 development and also regulates the placental supply of nutrients and the demand of nutrients by the
82 foetus (Constância *et al.* 2002).

83 To further understand the role epigenetic modifications such as DNA methylation have in
84 influencing phenotypes and their evolution, we need robust ways of manipulating them. Here we
85 consider the efficacy of the chemical 5-aza-2'-deoxycytidine (5-aza-dC) as a way of manipulating DNA
86 methylation, using the parasitoid wasp *Nasonia vitripennis*. The extent and nature of DNA methylation
87 across the *N. vitripennis* genome has recently been characterised (Wang et al. 2013; Pegoraro et al.
88 2016), and the presence in the genome of a full "methylation toolkit" has also been confirmed (Werren
89 et al. 2010). 5-aza-dC is well-known from the cancer literature as a hypomethylating agent, and is used
90 in the treatment of acute myeloid leukaemia and myelodysplastic syndrome (Seelan *et al.* 2018, Bryan
91 *et al.* 2011, Momparler 2012). In cancer patients, many genes that suppress leukemogenesis in
92 humans are silenced by aberrant DNA methylation. 5-aza-dC acts as a demethylating agent to
93 reactivate these genes (see below for details of action). In recent years, 5-aza-dC has been used across
94 a range of species as an experimental tool to explore how DNA methylation influences phenotypes
95 including diapause in *Nasonia vitripennis* (Pegoraro *et al.* 2016) and worker reproductive status in
96 bumble bees (Amarasinghe *et al.* 2014). Previously, we used 5-aza-dC to explore whether DNA
97 methylation was involved in the control of facultative sex allocation in *N. vitripennis* (Cook *et al.*
98 2015a).

99 Like all Hymenoptera, *N. vitripennis* is haplodiploid, with males arising from unfertilised
100 haploid eggs and females arising from fertilised diploid eggs. Females have putative control over the
101 sex of their offspring, releasing sperm to produce daughters or not to produce sons. Female *Nasonia*
102 allocate sex broadly in line with the predictions of Local Mate Competition theory (LMC: Hamilton
103 1967, Taylor and Bulmer 1980, Werren 1980, 1983, Shuker and West 2004, Burton-Chellew *et al.,*
104 2008; reviewed by West 2009). Females laying eggs alone on their blowfly pupae hosts produce very
105 female biased offspring sex ratios to minimise competition amongst her sons for mating opportunities.
106 As more females (foundresses) lay eggs on the same host(s) together, females produce less female-
107 biased sex ratios, as LMC is reduced and the fitness gained through the production of sons and

108 daughters moves towards equality (Hamilton 1967). When treated with 5-aza-dC, the facultative sex
109 allocation response in female *Nasonia* was maintained, suggesting that this plastic behaviour in and
110 of itself did not require certain patterns of methylation. Rather, the sex ratios (as proportion male)
111 produced by females were shifted slightly upwards, with more sons produced by treated females than
112 by controls. This subtle shift is as predicted by genomic conflict over sex allocation theory, which
113 requires some form of genomic imprinting (Wild and West 2008). Assuming that DNA methylation is
114 the mechanism by which parent-of-origin allelic information is carried in *Nasonia*, our data therefore
115 provided the first (albeit indirect) evidence for genomic conflict influencing sex allocation (Cook *et al*.
116 2015a).

117 Recently, Ellers *et al.* (2018) challenged the effects of 5-aza-dC as a general modifier of
118 methylation status, considering CpGs in ten genes in *Nasonia vitripennis*. Ellers *et al.* (2018) failed to
119 find changes in methylation status across any of the 155 CpGs in their study, questioning the action of
120 5-aza-dC in *Nasonia*, and hence of course the interpretation of the effects of 5-aza-dC in our earlier
121 study (Cook *et al.* 2015a). Fully negative results are of course very hard to interpret, but Ellers *et al.*
122 (2018) nonetheless raised valid concerns of the efficacy of the chemical. As noted above, levels of
123 methylation in insects is generally low, and a genome-wide study would provide greater resolution to
124 explore the effects of this chemical in an ecologically valuable model system. Here we consider the
125 effects of 5-aza-dC across the whole genome using a Bisulfite sequencing (BS-Seq) experiment. There
126 is evidence to suggest that the effects of 5-aza-dC are unlikely to be random across methylated CpGs
127 in a given genome. For instance, in mammals, genomic context can influence the extent to which 5-
128 aza-dC is effective as a demethylating agent. DNA sequence context, the distribution of transcription
129 factor binding sites, and chromatin structure are known to influence the DNA methylation patterns
130 produced by 5-aza-dC exposure (Ramos *et al.* 2015, Mossmann *et al.* 2010, Hagemann *et al.* 2011).
131 Although patterns of methylation differ among eukaryotes, the epigenetic modification itself is
132 chemically identical. Here, we verify that 5-aza-dC influences patterns of methylation across the
133 genome in *N. vitripennis*. Altered methylation patterns produced by 5-aza-dC exposure were non-
134 random, with enrichment in genes associated with transcription-factor activity and sequence specific
135 binding.

136

137

138 **Materials and Methods**

139 *Study species*

140 *Nasonia vitripennis* (Hymenoptera, Chalcicoidea) is a generalist parasitoid of large dipteran pupae
141 including species of Calliphoridae. Females oviposit between 20 and 50 eggs in an individual host, with
142 male offspring emerging just before females (after approximately 14 d at 25 °C: Whiting 1967). Males
143 are brachypterous and unable to fly, remaining close to the emergence site where they compete with
144 each other for emerging females, including their sisters. Females disperse after mating to locate new
145 hosts. The females used in this experiment were from the wild-type AsymC strain, originally isolated
146 in 1986 by curing the wild-type strain LabII of *Wolbachia*. Wasps have been maintained on *Calliphora*
147 *vomitoria* or *C. vicina* hosts at 25 °C, 16L:8D light conditions ever since, and AsymC is also the
148 reference genome strain (Werren *et al.* 2010). The quantitative genetic basis of sex ratio variation has
149 been repeatedly quantified (reviewed in Pannebakker et al. 2008, 2011) whilst whole-genome
150 transcription studies have confirmed that facultative sex allocation takes place without discernible
151 changes in gene expression (Cook et al. 2015b, 2018a).

152

153 *5-aza-2'-deoxycytidine*

154 We used 5-aza-dC (Sigma Aldrich Company Ltd., Gillingham, UK) at a non-lethal concentration (10 µM).
155 Widely used in epigenetic anticancer treatments (Seelan *et al.* 2018), 5-aza-dC is a nucleoside
156 analogue that works by inhibiting DNA methyltransferase activity (Piskala and Sorm 1964, Jones and
157 Taylor 1980, Momparler 1985, Christman 2002). When initially introduced to a cell, 5-aza-dC is
158 essentially inert but is then converted by processes inherent in the cell to its active form 5-aza-dCTP
159 (Seelan *et al.* 2018). This active form is readily incorporated into DNA during S-phase in the place of
160 cytosine. In the daughter strands arising from a replication event, DNA methyltransferase 1 (Dnmt1)
161 will attempt to restore the patterns of methylation that were present in the parent strand. However,
162 Dnmt1 becomes irreversibly bound to the 5-aza-dCTP residues in the daughter strands, resulting in a
163 diminishing pool of Dnmt1. Thus, administering 5-aza-dC results in a passive loss of DNA methylation;
164 the key effect of 5-aza-dC (for a fuller description, see Cook et al, 2018b).

165

166 *Experimental design*

167 To control for possible host and other maternal effects, experimental females were not drawn straight
168 from stock populations. Instead, two-day-old, mated, wild-type AsymC females were isolated from

169 the mass cultures into individual glass vials. Each female was provided with three hosts and allowed
170 to oviposit. Experimental females were drawn from the F1 generation, two days after emergence, one
171 female per "grandmother". We then employed a three by three factorial design whereby females
172 were allocated to one of three 5-aza-dC exposure regimes and were harvested post-feeding at three
173 time points. Exposure regimes were a) 20% sucrose for 24h, b) 20% sucrose supplemented with 10
174 µM 5-aza-dC for 24h and c) 20% sucrose supplemented with 10 µM 5-aza-dC for 48h. For each of the
175 three regimes, females were harvested at 0, 24 or 48h post-exposure. Throughout the experiment,
176 females were kept individually in glass vials and incubated at 25 °C, 16L :8D. Solutions (200 µl) were
177 provided in the lid of a 2 ml microcentrifuge tube. For individuals exposed to 5-aza-dC for 48 h, the
178 solutions were topped up by 100µl half-way through the exposure period to compensate for
179 evaporation. All solutions were stained green with food colouring such that a subset of 10 females
180 from each treatment combination could be dissected to check that feeding had occurred. At the point
181 of harvest, females were flash-frozen in liquid nitrogen and stored at -80 °C prior to DNA extraction.

182

183 *DNA extraction and bisulphite-conversion*

184 Prior to DNA extraction, individuals were removed from -80 °C storage and heads were quickly excised
185 from bodies using a sterile scalpel and a cryloyser to keep samples cold. Heads (and likewise bodies)
186 from the same treatment combination were randomly pooled into groups of 10 to give a maximum of
187 3 biological replicates for each treatment combination giving a total of 54 DNA samples. Briefly, tissue
188 was homogenised using a micropestle in 350 µl CTAB (Hexadecyltrimethylammonium Bromide) buffer
189 with subsequent overnight proteinase K digestion (400 µg) at 56 °C. Samples were cooled to room
190 temperature prior to RNase A digestion for 1 hour at 37 °C. A chloroform: isoamyl-alcohol wash (300
191 µl) was performed twice before ethanol/sodium acetate precipitation of the DNA. DNA purity was
192 analysed using the Nanodrop spectrophotometer with all samples having a 260/280 ratio of ≥1.8. DNA
193 integrity was confirmed using agarose gel electrophoresis.

194

195 *Bisulphite sequencing*

196 Bisulphite conversion, library preparation and sequencing were carried out by Edinburgh Genomics.
197 Briefly, bisulphite conversion was carried out using the EZ DNA Methylation Kit (Zymo Research, Irvine
198 USA), designed to reduce template degradation during the harsh bisulphite treatment whilst ensuring
199 complete conversion of the DNA. Library preparation was carried out using the TruSeq DNA
200 methylation Kit which takes the single-stranded DNA that results from bisulphite treatment and

converts it into an Illumina sequencing library. Sequencing of bisulphite-treated DNA was carried out on the Illumina HiSeq 2500 in high output mode with 125bp paired-end reads. Library preparation failed for three of the samples due to low levels of input DNA, leaving only two biological replicates for three out of 18 treatment combinations and a total of 51 Bisulphite-sequenced samples (Table 1).

Raw data processing

Initial data processing was carried out by Edinburgh Genomics. Briefly, reads were trimmed using Cutadapt version 1.12 removing any adapter sequences and for quality using a threshold of 30. After trimming reads were required to have a minimum length of 50 bp. *Nasonia vitripennis* genome version 2.1 and the associated annotation, available for download at Ensembl (http://metazoa.ensembl.org/Nasonia_vitripennis/Info/Index), were used as a reference. Reads were aligned to the reference genome using Bismark (version 0.16) with the parameter "bowtie2". PCR duplicates were removed from the resulting BAM files using samtools with the parameter "view -F 1024". Data for each CpG was then extracted from the BAM files using the Bioconductor package methylKit (version 0.99.3) with R (version 3.3.1). The methylation status of each CpG for all samples was calculated using the function "processBismarkAln" using the default parameters, except "min.cov=1 no;ap=TRUE", which excludes CpG's not covered by any reads across all samples and does not count CpGs covered by overlapping ends of a paired read as covered twice. The 1% of bases with the highest coverage were removed using the function "filterByCoverage" to eliminate potential PCR bias. After initial processing as outlined above, the raw data comprised percentage methylation at 4,150,376 CpG loci (no. of reads in which the CpG is methylated/total no. of reads covering the CpG) for each of 51 samples. CpG read counts were then summed by gene, (including 1Kb up and downstream) using gene annotation from Ensembl (version 2.1. available from: ftp://ftp.ensemblgenomes.org/pub/metazoa/release-40/gff3/nasonia_vitripennis). Around half (2,119,643) of the CpG's in the full dataset fell within genic regions with 14,765/17,279 genes (85%) having at least one CpG present.

Statistical analysis

The overall structure of our dataset in terms of the proportion of methylated reads per gene and per loci was visualised with two complementary analyses: a principle component analysis (PCA), which allows samples to be visualised in relation to the two major axes of variation in our experiment, and a T-distributed Stochastic Neighbour Embedding (t-SNE) analysis. t-SNE is a machine-learning algorithm

233 for data visualisation well-suited for high-dimensional datasets (van der Maaten and Hinton 2008).
234 The PCA was conducted using the R base function prcomp with default values. To perform the t-SNE
235 analysis we used the R package Rtsne (Krijthe 2015), with default values and perplexity set to 15. For
236 the PCA and t-SNE analyses any genes or loci with 0 coverage in any one sample were discarded prior
237 to analysis.

238 To examine the effect of 5-aza-dC exposure on individual genes, we modelled the number of
239 methylated and unmethylated reads from a gene as a generalized linear model (GLM) with a binomial
240 distribution with the following terms: exposure, tissue-type, collection time and their interactions.
241 Genes with 0 coverage for all samples in any particular treatment group were excluded from our
242 analysis (leaving a total of 11582 genes). Significance of each term was determined by a likelihood
243 ratio test, which was corrected for multiple testing using Benjamini and Hochberg's algorithm
244 (Benjamini and Hochberg 1995), with statistical significance set to 5%. Coefficients (log odds of
245 methylation) for each term were extracted from the full model. All statistical analyses were conducted
246 in R (v. 3.4.1; R Core Team 2017).

247

248 *Functional enrichment*

249 *N. vitripennis* GO-terms were downloaded from "hymenopteramine"
250 (http://hymenopteragenome.org/hymenopteramine/begin.do) on 27th September 2018. In addition,
251 we used hymenopteramine to firstly obtain *D. melanogaster* orthologs for each *N. vitripennis* gene,
252 from which we could then obtain the associated *D. melanogaster* GO terms. We used both of these
253 GO-term sets for enrichment analyses (below). Some *N. vitripennis* genes had multiple *D.*
254 *melanogaster* orthologs. In these cases, all GO-terms from each *D. melanogaster* ortholog were
255 combined together. To ensure this did not bias our results we repeated our analyses, but instead of
256 combining GO-terms we kept only GO-terms that were shared between all orthologs.

257 To examine which processes are most affected by 5-aza-dC exposure, we conducted gene set
258 enrichment analyses using the R package TopGO (v. 2.28.0) (Rahnenfuhrer 2016) using the elim
259 algorithm to account for the GO topology. Gene set enrichment analyses identify enriched GO terms
260 in a threshold-free way, by finding GO-terms that are overrepresented at the top of a ranked list of
261 genes. Here we ranked genes by FDR for the exposure effect. GO terms were considered to be
262 significantly enriched when $p < 0.05$.

263

264 **Results**

265 Median coverage across the whole experiment (N = 51 BS-seq libraries) was 916X, but coverage was
266 heterogeneous between loci (Figure 1) and samples (Table S1). Mean read coverage per sample
267 ranged from 2.7X to 49.4X. Four samples had less than 10X mean coverage per CpG, however, these
268 samples were distributed across treatment groups.

269 Treatment with 5-aza-dC had wide-ranging effects on DNA methylation across the genome.
270 First, samples clustered strongly by exposure regime at the gene-level in both the t-SNE visualisation
271 (Figure 2) and in the principle component analyses (Figure 3), indicating that 5-aza-dC has a strong
272 effect on the methylation of genes. Both analyses show the strong clustering at the gene loci, and that
273 wasps exposed for 24 hours are more like controls than wasps treated for 48 hours. In addition, the
274 PCA analysis showed clear separation along PCA1 (which presumably is heavily associated with
275 exposure regime; Figure 3). Importantly though, this analysis also emphasises the extent of variation
276 in response across the three treatment regimes, which is smallest for the controls, and largest for the
277 48 hour-exposed wasps (note the spread along PCA axis 2 in Figure 3). In addition, both cluster
278 analyses placed the controls as some way intermediate between the two 5-aza-dC exposed
279 treatments, which we consider further in the next two paragraphs. Both analyses also show that
280 exposure regime dominated the clustering, over and above of any effects of either tissue-type or
281 collection time, suggesting that this is the key effect. When we consider clustering at the level of
282 individual CpG loci, clustering is much less obvious, using both t-SNE and PCA. However, the control
283 treatment appears to cluster within the others in the t-SNE analyses, and the 24 hour and 48 hour
284 treatments do separate to some extent, e.g. with the 24h samples more the bottom and left of Figure
285 S1 and the 48h samples more to the top and right (see Figures S1 and S2). The differences between
286 the individual CpG and gene-level cluster analyses are likely to be due to a combination of low
287 coverage and a greater stochasticity of methylation at the individual CpG level (see discussion).

288 Second, a more in-depth examination of the data using a generalised linear modelling (GLM)
289 approach to examine differential methylation at the gene level found that 5-aza-dC exposure had a
290 significant effect on the methylation status of the majority of genes examined (exposure treatment
291 was significant as a main effect at an FDR level of 0.05 for 8,556 out of 11,582 genes, or 73.9% of
292 genes). Moreover, both tissue-type and collection time also influenced methylation status across a
293 large number of genes (significant as main effects in the GLMs for 3483 and 3134 genes respectively).
294 Significant interactions between factors were also present for the majority of genes (exposure regime
295 * collection time = 10,547 genes, exposure regime * tissue type = 9,042 genes, collection time * tissue
296 type = 5,011 genes, and exposure * collection time * tissue type = 10,234 genes), showing widespread

297 evidence for significant interplay between these factors. These results emphasise the context-
298 dependent nature of the action of 5-aza-dC. Full results from the GLMs are presented in Table S2.

299 Interestingly, it appears that methylation tends to increase initially in response to exposure to
300 5-aza-dC, an effect that is noticeable after 24h of exposure, before decreasing after 48h of exposure
301 (Figure 4A). This means that how long individuals are exposed to the chemical may well influence what
302 kind of methylation changes are observed; in other words, both hyper- and hypo-methylation can
303 result from 5-aza-dC treatment. In terms of when wasps were collected after exposure, 3134 genes
304 displayed significantly decreased methylation in association with collection time, with a more
305 pronounced decrease in methylation in samples collected 48h post-exposure relative to 24h post-
306 exposure (when compared to 0h post-exposure; Figure 4B). This suggests that the effect of 5-aza-dC
307 persists at least 48h after exposure has stopped. Finally, 3483 genes displayed significantly altered
308 methylation patterns in association with tissue type (Figure 4C), with lower methylation overall in the
309 head. However, we re-iterate that these main effects need to be contextualised within the pattern of
310 widespread interactions seen in the data, with 88% of genes exhibiting a significant second-order
311 interaction between exposure, tissue type and collection time.

312 Gene-set enrichment analysis, using *N. vitripennis* GO terms, when genes were ranked by
313 their response to exposure regime is presented in Table S3. Broadly, we identified 129 significantly
314 enriched GO terms, 52 of these terms fell under the "Biological process" (BP) category, 16 under
315 "Cellular component" (CC) and 61 under "Molecular function" (MF). The top three enriched GO
316 terms were "sequence-specific DNA binding" (GO:43565), "regulation of transcription, DNA
317 templated" (GO:6355) and "transcription factor activity, sequence-specific DNA binding" (GO:
318 43565) (Table S3). When the analysis was repeated using GO terms for the *D. melanogaster*
319 orthologs, a higher number of significantly enriched GO terms was obtained (686; 526 BP, CC 59, MF
320 89) of significantly enriched GO terms was obtained (Table S4). The top three GO terms for the *D.*
321 *melanogaster* analysis were "nucleus" (GO:0005634), "positive regulation of transcription from RNA
322 polymerase II promoter" (GO:45944), and, making a reappearance, "sequence-specific DNA binding"
323 (GO:43565). This analysis combined GO-terms when *N. vitripennis* genes had multiple *D.*
324 *melanogaster* orthologs, which could lead to bias, however, we retrieve very similar enriched terms
325 when we used only the GO terms that were shared between the multiple orthologs (Table S5),
326 suggesting any bias caused by this approach is small.

327

328

329 **Discussion**

330 A genome-wide analysis of CpGs across more than 11,000 genes, using 51 biological replicates across
331 our treatment combinations, has shown that the demethylating agent 5-aza-dC has very widespread
332 effects on patterns of DNA methylation in the *Nasonia vitripennis* genome. The methylome has been
333 characterised in *Nasonia vitripennis* by Wang et al (2013) and Pegoraro et al (2016), and in common
334 with other insect species, DNA methylation appears to be primarily associated with gene bodies
335 (including intron-exon boundaries), and a rather low proportion of CpGs are methylated. As
336 mentioned above, previous work that has attempted to confirm the effects of 5-aza-dC on DNA
337 methylation have had mixed results. Pegoraro et al (2016) confirmed the effects of 5-aza-dC across
338 five genes (four via a qPCR MethylQuant assay and one via cloning a fragment of the *msn* gene from
339 treated versus untreated individuals), whilst Ellers et al (2018) have recently failed to show any
340 effects of 5-aza-dC on the methylation status of 155 CpGs across ten genes (nine of which provided
341 libraries; they used high-throughput bisulfite amplicon sequencing: BSAS; see Redshaw et al. 2014
342 for a discussion of different methylation quantification assays). Our larger-scale study has both
343 confirmed the action of 5-aza-dC as a disruptor of methylation in *Nasonia*, but also suggested why it
344 might be so difficult to pin down those effects. In particular, we find that whilst we see statistically
345 significant main effects of exposure regime on DNA methylation (which includes differences in
346 length of exposure), we also see effects of tissue type (albeit at the crude level of heads versus
347 bodies), and collection time. Moreover, we also see many statistically significant interactions
348 between these factors; all told, the action of 5-aza-dC is likely to be context-dependent, both in
349 terms of the sequences more likely to be targeted but also in terms of the experimental details,
350 including target tissues. Nonetheless, our results strongly suggest that 5-aza-dC is not chemically
351 inert in terms of influencing DNA methylation in *Nasonia* – indeed, quite the opposite.

352 That the effects of 5-aza-dC are non-random in terms of the genes it influences are further
353 confirmed by our gene ontology enrichment analyses. We see enrichment for genes associated with
354 DNA binding, regulation of transcription, and transcription factor activity. These processes are of
355 interest given that we might expect the regulation of transcription to be a crucial component of the
356 epigenetic role of DNA methylation, given what we know from both mammals and insects, even
357 though the targets (throughout the genome versus exons) and outcomes (transcriptional silencing
358 versus stable gene expression) are typically different in the two taxa. Our analysis certainly confirms
359 the sequence-specific action of 5-aza-dC seen in other species.

360 A further important effect we see is that 5-aza-dC exposure can result in hyper-methylation
361 as well as hypo-methylation. This means that the characterisation of 5-aza-dC as a strictly

362 demethylating agent may be somewhat misleading. The complexity of outcomes of 5-aza-dC in
363 mammals, in particular in terms of trying to understand its action for cancer treatments, is becoming
364 increasingly well-appreciated (e.g. Seelan et al 2018). As such, our data confirm in an insect what is
365 being seen in mammalian cells. We are not well placed to speculate on the mechanism(s) by which
366 5-aza-dC may lead to hyper-methylation (see instead Seelan et al. 2018), even though its
367 hypomethylating effect is well characterised chemically (see Introduction). However, if 5-aza-dC
368 disrupts the regulation of the DNA methylation machinery itself, for instance in terms of how the
369 Dnmt genes are expressed and act, then the downstream outcomes could be both hyper- and hypo-
370 methylation. Perhaps related to this, in their study of DNA methylation and the control of the
371 diapause response in *Nasonia*, Pegoraro et al (2016) used both 5-aza-dC and RNA interference
372 (RNAi) to manipulate methylation, in the latter case disrupting DNA methylation by knocking down
373 both Dnmt1a and Dnmt3. Whilst both 5-aza-dC and the RNAi constructs influenced how diapause
374 was influenced by exposure to short or long day length, they did so in different ways. Whilst both
375 manipulations abolished the day-length response, the RNAi of Dnmt1a and Dnmt3 resulted in a
376 change in how wasps responded to long-day length (increasing the diapause response), whilst 5-aza-
377 dC led to a change on response to short-day length (decreasing the diapause response). As such,
378 different manipulations of DNA methylation may lead to different phenotypic outcomes, depending
379 on the regulatory causes and consequences of the methylation pattern itself. More generally then, it
380 is clear that there is much that remains to be elucidated in how 5-aza-dC shapes methylation
381 patterns, in both insects and mammals (Seelan et al. 2018).

382 On a more technical note, perhaps unsurprisingly we found more significant genes in the GO
383 enrichment analysis using the GO terms for *Drosophila melanogaster* orthologs. This is perhaps
384 because of the more detailed gene ontology information available for the *D. melanogaster* genome.
385 Nonetheless, the two analyses provide a similar story in terms of GO enrichment. In addition, the
386 majority of our analyses were undertaken at the level of the gene. Perhaps given the variation in
387 coverage at the level of individual CpGs, the clustering analyses were not by any means as clear cut
388 as the gene-level results. However, in the t-SNE analysis (Figure S1), the control replicates cluster
389 together, albeit within the broader distribution of the other two exposure treatment groups. The
390 effect of genomic and sequence context may be particularly relevant here: our analysis at the level
391 of the gene (see Methods for details) brings together CpGs that share a lot of context, by virtue of
392 being associated with a given genetic locus. The differences in the clustering at the individual CpG
393 versus gene levels may therefore represent the fact that the action of 5-aza-dC does appear to be
394 non-random, and that we gain clarity when we consider closely associated CpGs whose patterns of

395 methylation, and responses to 5-aza-dC, are more similar than randomly chosen CpGs across the
396 genome.

397 In conclusion, our results suggest that 5-aza-dC remains a potential tool for evolutionary
398 ecologists wishing to explore how DNA methylation influences phenotypic expression. As with all
399 genomic approaches however, no one approach gives all the answers. 5-aza-dC may prove a very
400 useful, if rather blunt, tool for first assessing the range of phenotypes that are influenced by
401 methylation status, whilst no doubt more targeted knock-downs and genetic manipulations will be
402 needed to link phenotypes to causal CpGs.

403

404 **Acknowledgments**

405 We are very grateful to Mark Lammers and Jacintha Ellers for their collegiate and constructive
406 engagement with our previous DNA methylation work. We are also grateful to Matthew Arno and
407 the Edinburgh Genomics team for their sequencing and downstream support. This work was
408 initiated as part of NERC grant NE/J024481/1.

409

410

411 **References**

412 Amarasinghe HE, Clayton CI and Mallon EB (2014) Methylation and worker reproduction in the
413 bumble-bee (*Bombus terrestris*). *Proceedings of the Royal Society B: Biological Sciences*
414 281(1780): 20132502.

415 Ashbrook DG, Mulligan MK and Williams RW (2018) Post-genomic behavioural genetics: From
416 revolution to routine. *Genes, Brain and Behaviour* 17: e12441.

417 Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: A practical and powerful approach
418 to multiple testing. J. R. Stat. Soc. Series B Stat. Methodol. 57:289–300.

419 Bewick AJ, Vogel KJ, Moore AJ and Schmitz RJ (2017) Evolution of DNA methylation across Insects.
420 *Molecular Biology and Evolution* 34(3), 654-665.

421 Biémont C (2010) From genotype to phenotype. What do epigenetics and epigenomics tell us?
422 *Heredity* 105(1), 1-3.

423 Bonduriansky R and Day T (2009) Nongenetic inheritance and its evolutionary implications. *Annual*
424 *Review of Ecology, Evolution and Systematics* 40(1), 103-125.

425 Bonduriansky R and Day T (2018) *Extended Heredity*. Princeton University Press: Princeton.

426 Bryan J, Kantarjian H, Garcia-Manero G and Jabbour E (2011) Pharmacokinetic evaluation of decitabine
427 for the treatment of leukemia. *Expert Opinion on Drug Metabolism and Toxicology* 7, 661-672.

428 Burggren W (2016) Epigenetic Inheritance and its role in evolutionary biology: re-evaluation and new
429 perspectives. *Biology* 5(2): 24.

430 Burton-Chellew MN, Koevoets T, Grillenberger BK, Sykes EM, Underwood SL *et al.* (2008) Facultative
431 sex ratio adjustment in natural populations of wasps: cues of local mate competition. *The*
432 *American Naturalist* 172(3), 393-404.

433 Christman JK (2002) 5-azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation:
434 mechanistic studies and their implications for cancer therapy. *Oncogene* 21(35), 5483-5495.

435 Constancia M, Hemberger M, Hughes J, Dean W, Fergusin-Smith A *et al.* (2002) Placental-specific IGF-
436 II is a major modulator of placental and fetal growth. *Nature* 417(6892), 945-948.

437 Cook N, Pannebakker BA, Tauber E and Shuker DM (2015a) DNA methylation and sex allocation in the
438 parasitoid wasp Nasonia vitripennis. *The American Naturalist* 186(4), 513-518.

439 Cook N, Trivedi U, Pannebakker BA, Blaxter M, Ritchie MG, Tauber E, Sneddon T and Shuker DM
440 (2015b) Oviposition but not sex allocation is associated with transcriptomic changes in
441 females of the parasitoid wasp *Nasonia vitripennis*. *G3: Genes, Genomes, Genetics*, 5, 2885-
442 2892.

443 Cook N, Boulton RA, Green J, Trivedi U, Tauber E, Pannebakker BA, Ritchie MG and Shuker DM (2018a)
444 Differential gene expression is not required for facultative sex allocation: a transcriptome
445 analysis of brain tissue in the parasitoid wasp *Nasonia vitripennis*. *Royal Society Open Science*,
446 5, 171718

447 Cook N, Parker DJ, Tauber E, Pannebakker BA and Shuker DM (2018b) Validating the demethylating
448 effects of 5-aza-2'-deoxycytidine in insects requires a whole-genome approach (A reply to
449 Ellers *et al.*) *The American Naturalist*, *in review*.

450 Ellers J, Visser M, Mariën J, Kraaijeveld K and Lammers M (2018) The importance of validating the
451 demethylating effect of 5-aza-dC in model species. Under Review.

452 Feng S, Cokus SJ, Zhang X, Chen P, Bostick M et al. (2010) Conservation and divergence of methylation
453 patterning in plants and animals. *Proceedings of the National Academy of Sciences of the USA*
454 107(19), 8689-8694.

455 Glastad KM, Hunt BG, Yi SV and Goodisman MAD (2011) DNA methylation in insects: on the brink of
456 the epigenomic era.

457 Hagemann S, Heil O, Lyko F and Brueckner B (2011) Azacytidine and decitabine induce gene-specific
458 and non-random DNA demethylation in human cancer cell lines. *PLOS ONE* 6(3), e17388.

459 Haig D (1997) Parental antagonism, relatedness asymmetries and genomic imprinting. *Proceedings of*
460 *the Royal Society B: Biological Sciences* 264(1388), 1657-1662.

461 Haig D (2002) Genomic imprinting and kinship. Rutgers University Press, New Brunswick, NJ.

462 Hamilton WD (1967) Extraordinary sex ratios. *Science* 156(3774), 477-488.

463 Jones PA and Taylor SM (1980) Cellular differentiation, cytidine analogs and DNA methylation. *Cell* 20,
464 85-93.

465 Krijthe JH (2015) Rtsne: T-Distributed Stochastic Neighbor Embedding using a Barnes-Hut
466 Implementation, URL: https://github.com/jkrijthe/Rtsne

467 Kucharski R, Maleszka J, Foret S and Maleszka R (2008) Nutritional control of reproductive status in
468 honeybees via DNA methylation. *Science* 319, 1827-1830.

469 Li Y, Zhang L, Yi Y, Hu W, Guo, Y, Zeng Z, Huang Z and Wang Z (2017) Genome-wide DNA methylation
470 changes associated with olfactory learning and memory in *Apis mellifera*. *Scientific Reports*
471 7(1): 17017.

472 Libbrecht R, Oxley PR, Keller L, Kronauer DJ (2016) Robust DNA methylation in the clonal raider ant
473 brain. *Current Biology* 26(3), 391-395.

474 Lind MI and Spagopoulou F (2018) Evolutionary consequences of epigenetic inheritance. *Heredity*
475 121(3), 205-209.

476 Lyko F, Foret S, Kucharski R, Wolf S, Falckenhayn C and Maleszka R (2010) The honey bee epigenomes:
477 differential metjhylation of brain DNA in queens and workers. *PLOS Biology* 8(11): e1000506.

478 Lyko F (2018) The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nature*
479 *Reviews Genetics* 19(2),81-92.

480 Momparler RL (1985) Molecular, cellular and animal pharmacology of 5-aza-2'-deoxycytidine.
481 *Pharmacology and Therapeutics* 30(3), 287-299.

482 Momparler RL (2012) A perspective on the comparative antileukemic activity of 5-aza-2'-
483 deoxycytidine (Decitabine) and 5-azacytidine (Vidaza). *Pharmaceuticals* 5(8), 875-881.

484 Mossman D, Kim K and Scott RJ (2010) Demethylation by 5-aza-2'-deoxycytidine in colorectal cancer
485 cells targets genomic DNA whilst promoter CpG island methylation persists. *BMC Cancer*
486 10:366.

487 Pannebakker BA, Halligan DL, Reynolds KT, Ballantyne GA, Shuker DM, Barton N and West SA (2008)
488 Effects of spontaneous mutation accumulation on sex ratio traits. *Evolution*, 62: 1921-1935.

489 Pannebakker BA, Watt R, Knott SA, West SA and Shuker DM (2011) The quantitative genetic basis of
490 sex ratio variation in *Nasonia vitripennis*: a QTL study. *Journal of Evolutionary Biology*, 24, 12-
491 22.

492 Patten MM, Ross L, Curley JP, Queller DC, Bonduriansky R and Wolf JB (2014) The evolution of genomic
493 imprinting: theories, predictions and empirical tests. *Heredity* 113, 119-128.

494 Pegoraro M, Bafna A, Davies NJ, Shuker DM and Tauber E (2016) DNA methylation changes induced
495 by long and short photoperiods in Nasonia. *Genome Research* 26(2), 203-210.

496 Piskala A and Sorm F (1964) Nucleic acid components and their analogues. LI. Synthesis of 1-glycosyl
497 derivatives of 5-azauracil and 5-azacytosine. *Collection of Czechoslovak Chemical*
498 *Communications* 29(9), 2060-2076.

499 R Core Team (2017). R: A language and environment for statistical computing. R Foundation for
500 Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

501 Ramos M, Wijetunga NA, McLellan AS, Suzuki M and Greally JM (2015) DNA demethylation by 5-aza-
502 2'-deoxycytidine is imprinted, targeted to euchromatin and has limited transcriptional
503 consequences. *Epigenetics and Chromatin* 8:11.

504 Rahnenfuhrer AAJ (2016) *topGO: Enrichment Analysis for Gene Ontology*. R package version 2.28.0.

505 Redshaw N, Huggett JF, Taylor MS, Carole A Foy CA and Devonshire AS (2014) Quantification of
506 epigenetic biomarkers: an evaluation of established and emerging methods for DNA
507 methylation analysis. BMC Genomics 15:1174

508 Seelan RS, Mukhopadhyay P, Pisano MM and Greene RM (2018) Effects of 5-Aza-2'-deoxycytidine
509 (decitabine) on gene expression. *Drug Metabolism Reviews* 50(2), 193-207.

510 Shuker DM and West SA (2004) Information constraints and the precision of adaptation: sex ratio
511 manipulation in wasps. *Proceedings of the National Academy of Sciences of the USA* 101(28),
512 10363-10367.

513 Smith CR, Mutti NS, Jasper WC, Naidu A, Smith CD and Gadau J (2012) Patterns of DNA methylation in
514 development, division of labor and hybridisation in an ant with genetic caste determination.
515 *PLOS One* 7(8): e42433.

516 Taylor PD and Bulmer MG (1980) Local Mate Competition and the sex-ratio. *Journal of Theoretical*
517 *Biology* 86(3) 409-419.

518 van der Maaten LJP and Hinton GE (2008) Visualizing high-dimensional data using t-SNE. *Journal of*
519 *Machine Learning Research* 9, 2579-2605.Wang X, Wheeler D, Avery A, Rago A, Choi JH,
520 Colbourne JK, Clark AG, and Werren JH (2013) Function and evolution of DNA methylation in
521 *Nasonia vitripennis*. *PLoS Genetics*. 9(10):e1003872.

522 Werren JH (1980) Sex ratio adaptations to local mate competition in a parasitic wasp. *Science*
523 208(4448), 1157-1159.

524 Werren JH (1983) Sex ratio evolution under local mate competition in a parasitic wasp. *Evolution* 37(1),
525 116-124.

526 Werren JH, *et al* (2010) Functional and evolutionary insights from the genomes of three parasitoid
527 *Nasonia* species. *Science*, 327: 343-348.

528 West SA (2009) Sex allocation. Princeton University Press, Princeton, NJ.

529 Whiting AR (1967) The biology of the parasitic wasp *Mormoniella vitripennis* [=*Nasonia brevicornis*]
530 (Walker). *The Quarterly Review of Biology* 42(3), 333-406.

531 Wild G and West SA (2009) Genomic imprinting and sex allocation. *The American Naturalist* 173(1),
532 E1-14.

533 Zemach A, McDaniel IE, Silva P and Zilberman D (2010) Genome-wide evolutionary analysis of
534 eukaryotic DNA methylation. *Science* 328(5980), 916-919.

535

536 **Data and code availability**

537 Raw reads will be deposited in the SRA and accession codes will be made available upon publication.

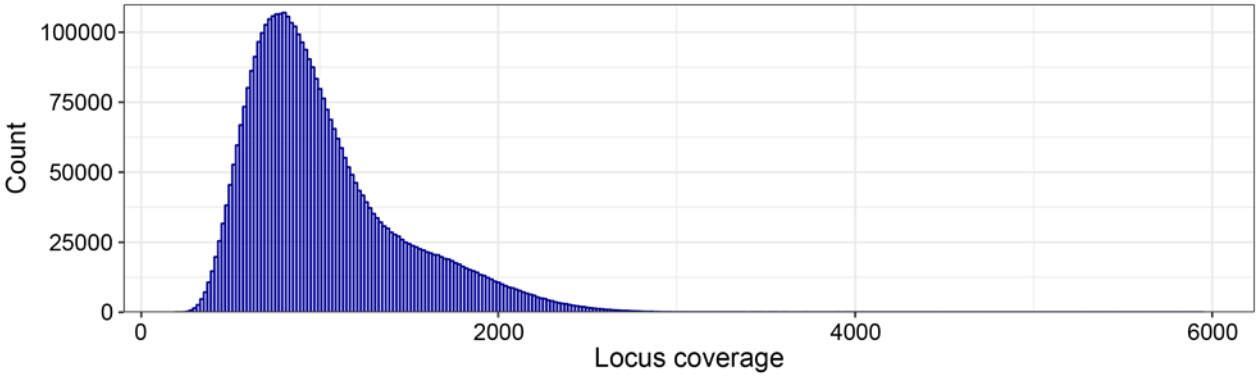
538 Scripts for the analyses in this paper are available at: https://github.com/DarrenJParker/5-Aza-

539 dC_methylation

540

541 **Tables and Figures**

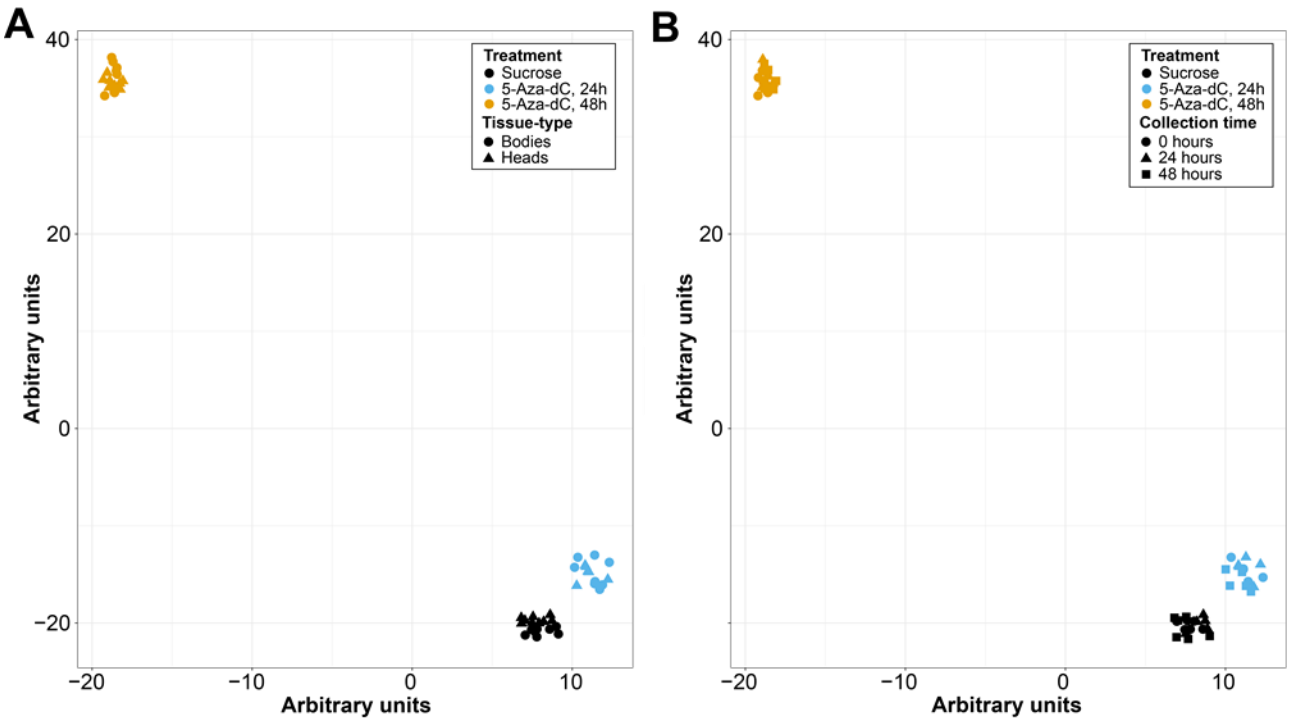
542



543

544 **Figure 1.** Read coverage by locus across all 51 BS-Seq libraries.

545



546

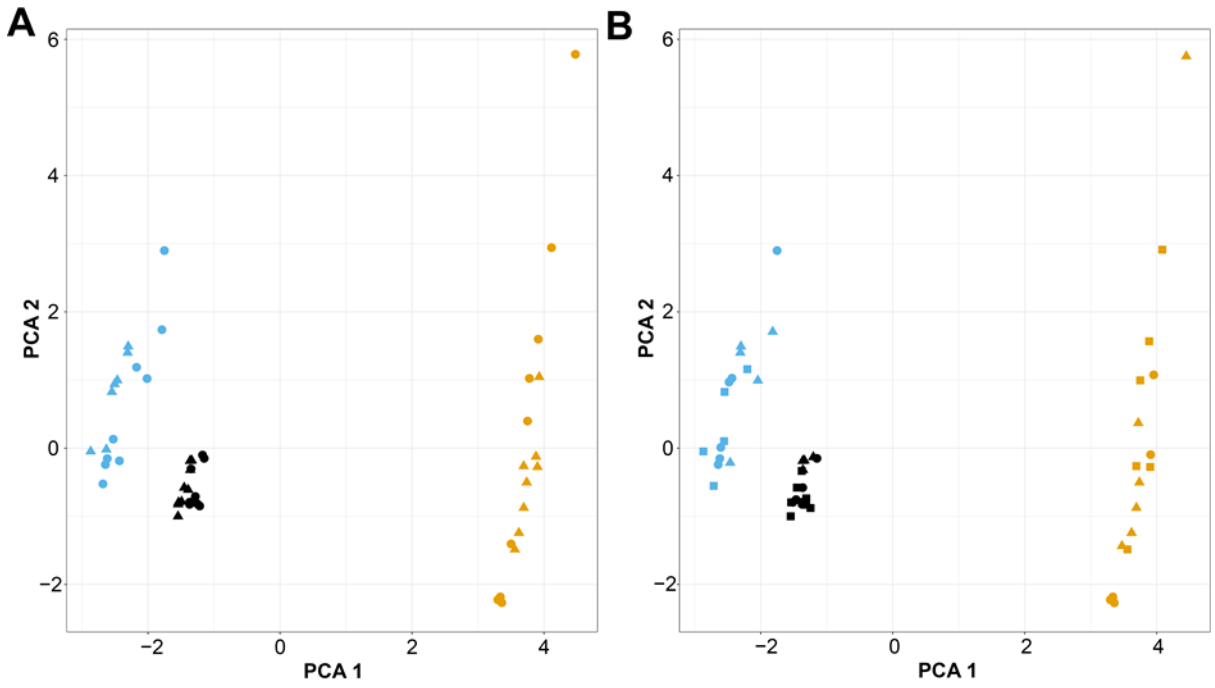
547 **Figure 2.** t-SNE of the proportion of methylated reads per gene, showing the effects of (A) Exposure
548 regime and tissue type and (B) Exposure regime and harvest time.

549

550

551

552

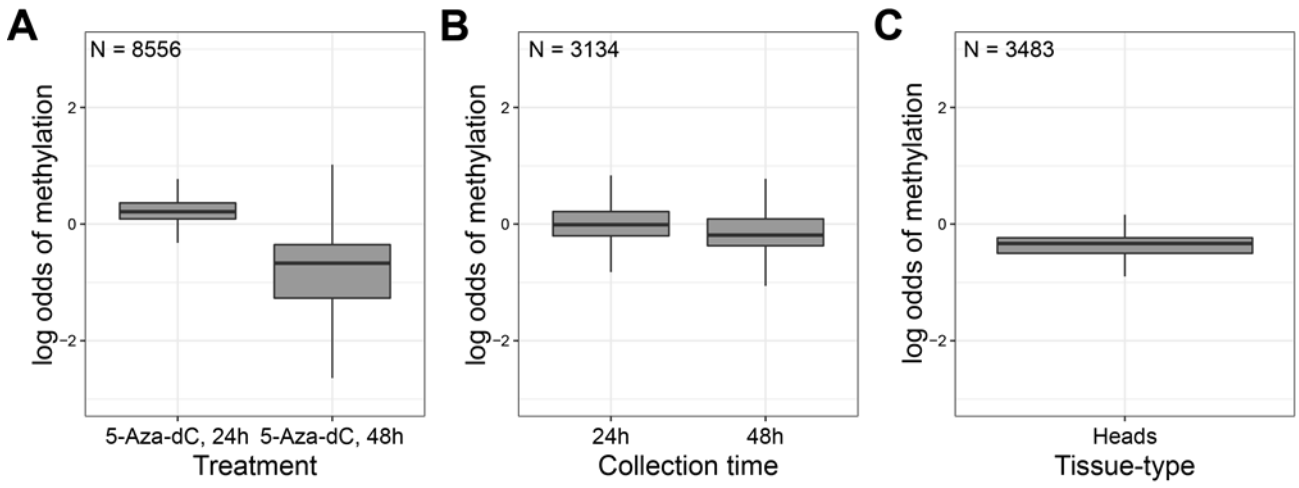


553

554 **Figure 3.** PCA of the proportion of methylated reads per gene, showing the effects of (A) Exposure
555 regime and tissue type and (B) Exposure regime and harvest time.

556

557



558

559 **Figure 4.** A significant effect on gene methylation status (displayed as log odds of methylation) was
560 noted in response to (A) 5-aza-dC exposure regime, (B) harvest time after exposure and (C) tissue-
561 type. N= number of genes where significantly differentially methylated CpGs in association with each
562 factor.

563

564 **Table 1.** Experimental design. Female wasps were exposed to 1 of 3 5-aza-dC exposure regimes
565 including a control. Females were then harvested at 1 of 3 timepoints after the exposure period
566 heads were excised from bodies to give a degree of tissue-specificity to the analysis. Three biological
567 replicates were available for almost all treatment combinations with each biological replicate
568 consisting of tissue from three *N. vitripennis* females.

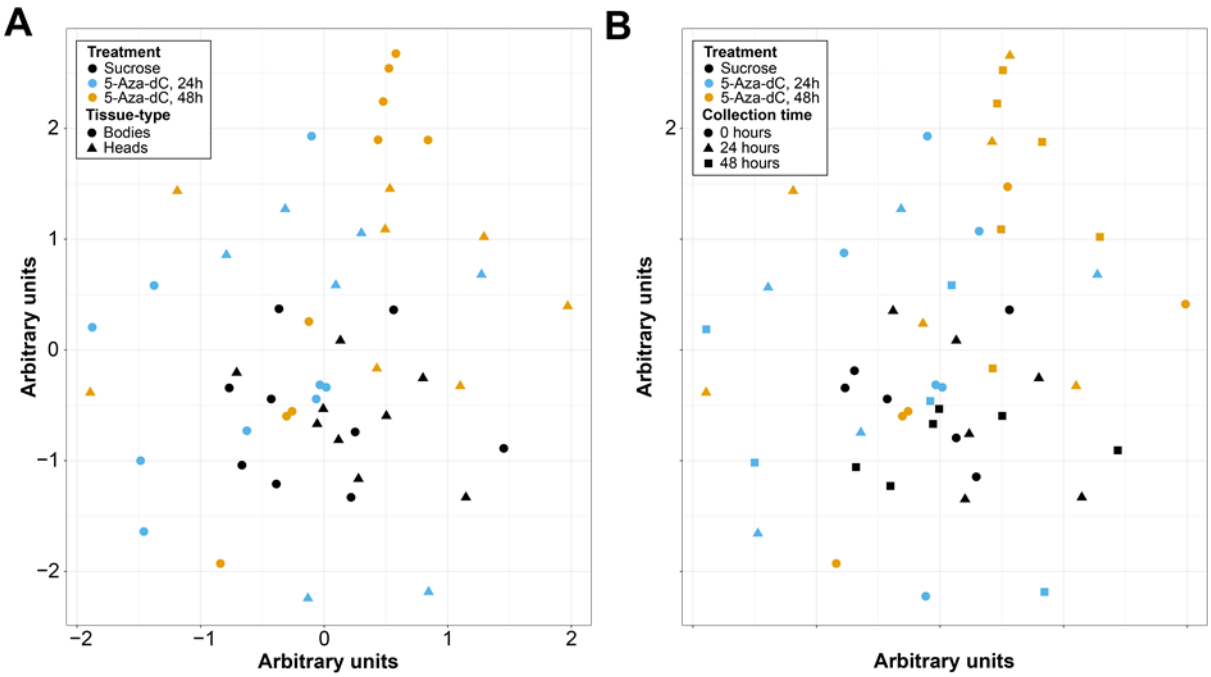
Treatment Combination	5-aza-dC exposure regime	Harvest time	Tissue type	No. replicates (1 x rep. = 1 pool of 10)
1	20% sucrose for 24 h	0h	Bodies	3
2	20% sucrose for 24 h	0h	Heads	3
3	20% sucrose for 24 h	24h	Bodies	3
4	20% sucrose for 24 h	24h	Heads	3
5	20% sucrose for 24 h	48h	Bodies	3
6	20% sucrose for 24 h	48h	Heads	3
7	10µM 5-aza-dC for 24 h	0h	Bodies	3
8	10µM 5-aza-dC for 24 h	0h	Heads	3
9	10µM 5-aza-dC for 24 h	24h	Bodies	3
10	10µM 5-aza-dC for 24 h	24h	Heads	2
11	10µM 5-aza-dC for 24 h	48h	Bodies	3
12	10µM 5-aza-dC for 24 h	48h	Heads	2
13	10µM 5-aza-dC for 48 h	0h	Bodies	3
14	10µM 5-aza-dC for 48 h	0h	Heads	2
15	10µM 5-aza-dC for 48 h	24h	Bodies	3
16	10µM 5-aza-dC for 48 h	24h	Heads	3
17	10µM 5-aza-dC for 48 h	48h	Bodies	3
18	10µM 5-aza-dC for 48 h	48h	Heads	3

569

570

571 **Supplementary Figures**

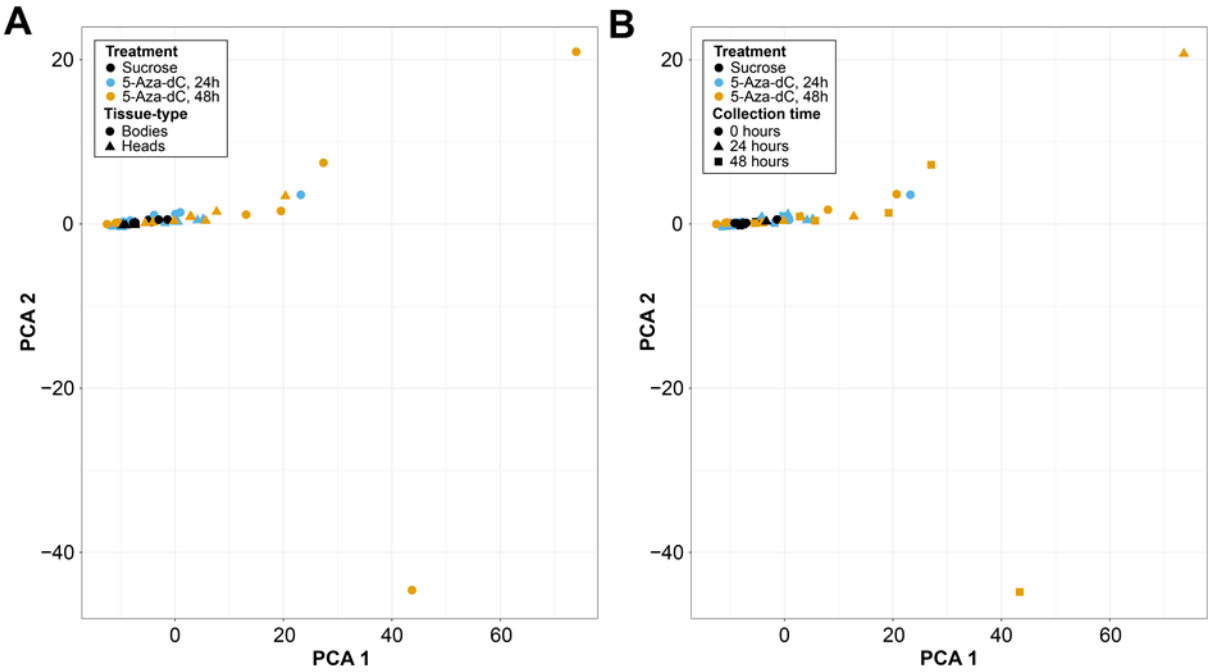
572



573

574 **Figure S1.** t-SNE of the proportion of methylated reads per CpG locus, showing the effects of (A)
575 Exposure regime and tissue type and (B) Exposure regime and harvest time.

576



577

578 **Figure S2.** PCA of the proportion of methylated reads per CpG locus, showing the effects of (A)
579 Exposure regime and tissue type and (B) Exposure regime and harvest time.

580

581 **Supplemental Table Information**

582

583 **Table S1.** Number of reads and coverage of each sample

584 **Table S2.** Output of coefficients and significance values from the GLM analysis

585 **Table S3.** Enriched GO terms for treatment using *Nasonia vitripennis* GO terms

586 **Table S4.** Enriched GO terms for treatment using *D. melanogaster* GO terms, when GO-terms from
587 multiple orthologs are combined together

588 **Table S5.** Enriched GO terms for treatment using *D. melanogaster* GO terms, when only GO-terms
589 which are the same are kept (when a *Nasonia vitripennis* gene has multiple *D. melanogaster*
590 orthologs)