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

Nicola Cook[1¶*], Darren J Parker[2,3¶], Frances Turner[4], Eran Tauber[5], Bart A Pannebakker[6] and David M Shuker[1]

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

*Corresponding author

E-mail: nc52@st-andrews.ac.uk

¶These authors contributed equally

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.

Author Summary

Shedding light on the mechanistic basis of phenotypes is a major aim in the field of evolutionary biology. If we understand how phenotypes are controlled at the molecular level, we can begin to understand how evolution has shaped that phenotype and conversely, how genetic architecture may constrain trait evolution. Epigenetic markers (such as DNA methylation) also influence phenotypic expression by regulating how and when genes are expressed. Recently, 5-aza-2'-deoxycytidine (5-aza-dC), a hypomethylating agent used in the treatment of certain cancers, has been used to explore the epigenetic regulation of traits of interest to evolutionary ecologists. Previously, we used 5-aza-dC to validate a role for DNA methylation in facultative sex allocation behaviour in the parasitoid wasp *Nasonia vitripennis*. However, the direct effects of the chemical were not examined at that point and its efficacy in insects was questioned. Here, we demonstrate that 5-aza-dC disrupts DNA methylation

on a genome-wide scale in a context- and sequence-specific manner and results in both hypo- and hyper-methylation. Our work demonstrates that 5-aza-dC has the potential to be repurposed as a tool for studying the role of DNA methylation in phenotypic expression.

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 [1, 2]. 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 [3]. 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 [4, 5]. By contrast, in insects, DNA methylation is concentrated in gene bodies and is associated with more stable patterns of gene expression [6]. 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 [6, 7]. 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 [5]. However, DNA methylation status influences how genes (and phenotypes) are expressed in many organisms, including reproductive vs non-reproductive worker bumblebees [8], caste differentiation [9, 10], and learning and memory [11] in honey bees, and behavioural division of labour in ants (*Pogonomyrmex barbatus* [12]; but see [13] 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 [14, 15]. Moreover, there has also been great interest in the extent to which epigenetic modifications may be inherited

92 across generations [14, 16, 17]. Transmission of epigenetic state allows transmission of parent-of-
93 origin information, or so-called genomic imprinting, setting the scene for genomic conflict [18, 19].
94 Genomic conflict arises when maternally- and paternally-inherited alleles have different evolutionary
95 optima for traits expressed in an individual [19, 20]. A classic example is the expression of maternally-
96 or paternally-inherited alleles during embryonic development in mammals: paternally-inherited
97 alleles are selected to extract more resources from the mother *in utero* compared to maternally-
98 inherited alleles [21]. This conflict is now known to be, at least in part, facilitated by parent-of-origin
99 specific gene expression of insulin-like growth factor II (Igf2) and Igf2r. In mammals, the Igf2 gene
100 promotes growth and cellular differentiation during development, and also regulates the placental
101 supply of nutrients and the demand of nutrients by the foetus [22].

102 To further understand the role epigenetic modifications such as DNA methylation have in
103 influencing phenotypes and their evolution, we need robust ways of manipulating them. Here we
104 consider the efficacy of the chemical 5-aza-2'-deoxycytidine (5-aza-dC) as a way of manipulating DNA
105 methylation, using the parasitoid wasp *Nasonia vitripennis*. The extent and nature of DNA methylation
106 across the *N. vitripennis* genome has recently been characterised [23, 24], and the presence in the
107 genome of a full "methylation toolkit" has also been confirmed [25]. 5-aza-dC is well-known from the
108 cancer literature as a hypomethylating agent, and is used in the treatment of acute myeloid leukaemia
109 and myelodysplastic syndrome [26-28]. In cancer patients, many genes that suppress leukemogenesis
110 in humans are silenced by aberrant DNA methylation. 5-aza-dC acts as a demethylating agent to
111 reactivate these genes (see below for details of action). In recent years, 5-aza-dC has been used across
112 a range of species as an experimental tool to explore how DNA methylation influences phenotypes
113 including diapause in *Nasonia vitripennis* [24] and worker reproductive status in bumble bees [8].
114 Previously, we used 5-aza-dC to explore whether DNA methylation was involved in the control of
115 facultative sex allocation in *N. vitripennis* [29].

116 Like all Hymenoptera, *N. vitripennis* is haplodiploid, with males arising from unfertilised
117 haploid eggs and females arising from fertilised diploid eggs. Females have putative control over the
118 sex of their offspring, releasing sperm to produce daughters or not to produce sons. Female *Nasonia*
119 allocate sex broadly in line with the predictions of Local Mate Competition theory [30-35; reviewed by
120 36]. Females laying eggs alone on their blowfly pupae hosts produce very female-biased offspring sex
121 ratios to minimise competition amongst her sons for mating opportunities. As more females
122 (foundresses) lay eggs on the same host(s) together, females produce less female-biased sex ratios,
123 as LMC is reduced and the fitness gained through the production of sons and daughters moves
124 towards equality [30]. When treated with 5-aza-dC, the facultative sex allocation response in female
125 *Nasonia* was maintained, suggesting that this plastic behaviour in and of itself did not require certain

126 patterns of methylation. Rather, the sex ratios (as proportion male) produced by females were shifted
127 slightly upwards, with more sons produced by treated females than by controls. This subtle shift is as
128 predicted by genomic conflict over sex allocation theory, which requires some form of genomic
129 imprinting [37]. Assuming that DNA methylation is the mechanism by which parent-of-origin allelic
130 information is carried in *Nasonia*, our data therefore provided the first (albeit indirect) evidence for
131 genomic conflict influencing sex allocation [29].

132 Recently, Ellers *et al.* [38] challenged the effects of 5-aza-dC as a general modifier of
133 methylation status, considering CpGs in ten genes in *Nasonia vitripennis*. Ellers *et al.* [38] failed to find
134 changes in methylation status across any of the 155 CpGs in their study, questioning the action of 5-
135 aza-dC in *Nasonia*, and hence of course the interpretation of the effects of 5-aza-dC in our earlier study
136 [29]. Fully negative results are of course very hard to interpret, but Ellers *et al.* [38] nonetheless raised
137 valid concerns of the efficacy of the chemical. As noted above, levels of methylation in insects is
138 generally low, and a genome-wide study would provide greater resolution to explore the effects of
139 this chemical in an ecologically valuable model system. Here we consider the effects of 5-aza-dC across
140 the whole genome using a Bisulfite sequencing (BS-Seq) experiment. There is evidence to suggest that
141 the effects of 5-aza-dC are unlikely to be random across methylated CpGs in a given genome. For
142 instance, in mammals, genomic context can influence the extent to which 5-aza-dC is effective as a
143 demethylating agent. DNA sequence context, the distribution of transcription factor binding sites, and
144 chromatin structure are known to influence the DNA methylation patterns produced by 5-aza-dC
145 exposure [39-41]. Although patterns of methylation differ among eukaryotes, the epigenetic
146 modification itself is chemically identical. Here, we verify that 5-aza-dC influences patterns of
147 methylation across the genome in *N. vitripennis*. Altered methylation patterns produced by 5-aza-dC
148 exposure were non-random, with enrichment in genes associated with transcription-factor activity
149 and sequence specific binding.

150

151

152 **Materials and Methods**

153 *Study species*

154 *Nasonia vitripennis* (Hymenoptera, Chalcicoidea) is a generalist parasitoid of large dipteran pupae
155 including species of Calliphoridae. Females oviposit between 20 and 50 eggs in an individual host, with
156 male offspring emerging just before females (after approximately 14 d at 25 °C: [42]). Males are
157 brachypterous and unable to fly, remaining close to the emergence site where they compete with
158 each other for emerging females, including their sisters. Females disperse after mating to locate new
159 hosts. The females used in this experiment were from the wild-type AsymC strain, originally isolated
160 in 1986 by curing the wild-type strain LabII of *Wolbachia*. Wasps have been maintained on *Calliphora*
161 *vomitoria* or *C. vicina* hosts at 25 °C, 16L:8D light conditions ever since, and AsymC is also the reference
162 genome strain [25]. The quantitative genetic basis of sex ratio variation has been repeatedly quantified
163 [43, 44], whilst whole-genome transcription studies have confirmed that facultative sex allocation
164 takes place without discernible changes in gene expression [45, 46].

165

166 *5-aza-2'-deoxycytidine*

167 We used 5-aza-dC (Sigma Aldrich Company Ltd., Gillingham, UK) at a non-lethal concentration (10 µM).
168 Widely used in epigenetic anticancer treatments [26], 5-aza-dC is a nucleoside analogue that works by
169 inhibiting DNA methyltransferase activity [47-50]. When initially introduced to a cell, 5-aza-dC is
170 essentially inert but is then converted by processes inherent in the cell to its active form 5-aza-dCTP
171 [26]. This active form is readily incorporated into DNA during S-phase of the cell cycle in the place of
172 cytosine. In the daughter strands arising from a replication event, DNA methyltransferase 1 (Dnmt1)
173 will attempt to restore the patterns of methylation that were present in the parent strand. However,
174 Dnmt1 becomes irreversibly bound to the 5-aza-dCTP residues in the daughter strands, resulting in a
175 diminishing pool of Dnmt1. Thus, administering 5-aza-dC results in a passive loss of DNA methylation,
176 the key effect of 5-aza-dC [for a fuller description, see 51].

177

178 *Experimental design*

179 To control for possible host and other maternal effects, experimental females were not drawn straight
180 from stock populations. Instead, two-day-old, mated, wild-type AsymC females were isolated from
181 the mass cultures into individual glass vials. Each female was provided with three hosts and allowed
182 to oviposit. Experimental females were drawn from the F1 generation, two days after emergence, one

183 female per “grandmother”. We then employed a three by three factorial design whereby females
184 were allocated to one of three 5-aza-dC exposure regimes and were harvested post-feeding at three
185 time points. Exposure regimes were a) 20% sucrose for 24h, b) 20% sucrose supplemented with 10
186 µM 5-aza-dC for 24h and c) 20% sucrose supplemented with 10 µM 5-aza-dC for 48h. For each of the
187 three regimes, females were harvested at 0, 24 or 48h post-exposure. Throughout the experiment,
188 females were kept individually in glass vials and incubated at 25 °C, 16L :8D. Solutions (200 µl) were
189 provided in the lid of a 2 ml microcentrifuge tube. For individuals exposed to 5-aza-dC for 48 h, the
190 solutions were topped up by 100µl half-way through the exposure period to compensate for
191 evaporation. All solutions were stained green with food colouring such that a subset of 10 females
192 from each treatment combination could be dissected to check that feeding had occurred. At the point
193 of harvest, females were flash-frozen in liquid nitrogen and stored at -80 °C prior to DNA extraction.

194

195 *DNA extraction and bisulphite-conversion*

196 Prior to DNA extraction, individuals were removed from -80 °C storage and heads were quickly excised
197 from bodies using a sterile scalpel and a cryloyser to keep samples cold. Heads (and likewise bodies)
198 from the same treatment combination were randomly pooled into groups of 10 to give a maximum of
199 3 biological replicates for each treatment combination giving a total of 54 DNA samples. Briefly, tissue
200 was homogenised using a micropestle in 350 µl CTAB (Hexadecyltrimethylammonium Bromide) buffer
201 with subsequent overnight proteinase K digestion (400 µg) at 56 °C. Samples were cooled to room
202 temperature prior to RNase A digestion for 1 hour at 37 °C. A chloroform: isoamyl-alcohol wash (300
203 µl) was performed twice before ethanol/sodium acetate precipitation of the DNA. DNA purity was
204 analysed using the Nanodrop spectrophotometer with all samples having a 260/280 ratio of ≥1.8. DNA
205 integrity was confirmed using agarose gel electrophoresis.

206

207 *Bisulphite sequencing*

208 Bisulphite conversion, library preparation and sequencing were carried out by Edinburgh Genomics.
209 Briefly, bisulphite conversion was carried out using the EZ DNA Methylation Kit (Zymo Research, Irvine
210 USA), designed to reduce template degradation during the harsh bisulphite treatment whilst ensuring
211 complete conversion of the DNA. Library preparation was carried out using the TruSeq DNA
212 methylation Kit which takes the single-stranded DNA that results from bisulphite treatment and
213 converts it into an Illumina sequencing library. Sequencing of bisulphite-treated DNA was carried out
214 on the Illumina HiSeq 2500 in high output mode with 125bp paired-end reads. Library preparation

215 failed for three of the samples due to low levels of input DNA, leaving only two biological replicates
216 for three out of 18 treatment combinations and a total of 51 Bisulphite-sequenced samples (Table 1).

217 **Table 1. Experimental design.** Female wasps were exposed to 1 of 3 5-aza-dC exposure regimes
218 including a control. Females were then harvested at 1 of 3 timepoints after the exposure period
219 heads were excised from bodies to give a degree of tissue-specificity to the analysis. Three biological
220 replicates were available for almost all treatment combinations with each biological replicate
221 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

222

223 *Raw data processing*

224 Initial data processing was carried out by Edinburgh Genomics. Briefly, reads were trimmed using
225 Cutadapt version 1.12 removing any adapter sequences and for quality using a threshold of 30. After
226 trimming reads were required to have a minimum length of 50 bp. *Nasonia vitripennis* genome version
227 2.1 and the associated annotation, available for download at Ensembl
228 (http://metazoa.ensembl.org/Nasonia_vitripennis/Info/Index), were used as a reference. Reads were
229 aligned to the reference genome using Bismark (version 0.16) with the parameter "bowtie2". PCR
230 duplicates were removed from the resulting BAM files using samtools with the parameter "view -F
231 1024". Data for each CpG was then extracted from the BAM files using the Bioconductor package
232 methylKit (version 0.99.3) with R (version 3.3.1; [52]). The methylation status of each CpG for all
233 samples was calculated using the function "processBismarkAln" using the default parameters, except
234 "min.cov=1 no;ap=TRUE", which excludes CpG's not covered by any reads across all samples and does

235 not count CpGs covered by overlapping ends of a paired read as covered twice. The 1% of bases with
236 the highest coverage were removed using the function "filterByCoverage" to eliminate potential PCR
237 bias. After initial processing as outlined above, the raw data comprised percentage methylation at
238 4,150,376 CpG loci (no. of reads in which the CpG is methylated/total no. of reads covering the CpG)
239 for each of 51 samples. CpG read counts were then summed by gene, (including 1Kb up and
240 downstream) using gene annotation from Ensembl (version 2.1. available from:
241 ftp://ftp.ensemblgenomes.org/pub/metazoa/release-40/gff3/nasonia_vitripennis). Around half
242 (2,119,643) of the CpG's in the full dataset fell within genic regions with 14,765/17,279 genes (85%)
243 having at least one CpG present.

244

245 *Statistical analysis*

246 The overall structure of our dataset in terms of the proportion of methylated reads per gene and per
247 loci was visualised with two complementary analyses: a principle component analysis (PCA), which
248 allows samples to be visualised in relation to the two major axes of variation in our experiment, and a
249 T-distributed Stochastic Neighbour Embedding (t-SNE) analysis. t-SNE is a machine-learning algorithm
250 for data visualisation well-suited for high-dimensional datasets [53]. The PCA was conducted using the
251 R base function prcomp with default values. To perform the t-SNE analysis we used the R package
252 Rtsne [54], with default values and perplexity set to 15. For the PCA and t-SNE analyses any genes or
253 loci with 0 coverage in any one sample were discarded prior to analysis.

254 To examine the effect of 5-aza-dC exposure on individual genes, we modelled the number of
255 methylated and unmethylated reads from a gene as a generalized linear model (GLM) with a binomial
256 distribution with the following terms: exposure, tissue-type, collection time and their interactions.
257 Genes with 0 coverage for all samples in any particular treatment group were excluded from our
258 analysis (leaving a total of 11582 genes). Significance of each term was determined by a likelihood
259 ratio test, which was corrected for multiple testing using Benjamini and Hochberg's algorithm [55],
260 with statistical significance set to 5%. Coefficients (log odds of methylation) for each term were
261 extracted from the full model. All statistical analyses were conducted in R (v. 3.4.1; [52]).

262

263 *Functional enrichment*

264 *N. vitripennis* GO-terms were downloaded from "hymenopteramine"
265 (http://hymenopteragenome.org/hymenopteramine/begin.do) on 27th September 2018. In addition,
266 we used hymenopteramine to firstly obtain *D. melanogaster* orthologs for each *N. vitripennis* gene,

267 from which we could then obtain the associated *D. melanogaster* GO terms. We used both of these
268 GO-term sets for enrichment analyses (below). Some *N. vitripennis* genes had multiple *D.*
269 *melanogaster* orthologs. In these cases, all GO-terms from each *D. melanogaster* ortholog were
270 combined together. To ensure this did not bias our results we repeated our analyses, but instead of
271 combining GO-terms we kept only GO-terms that were shared between all orthologs.

272 To examine which processes are most affected by 5-aza-dC exposure, we conducted gene set
273 enrichment analyses using the R package TopGO (v. 2.28.0; [56]) using the elim algorithm to account
274 for the GO topology. Gene set enrichment analyses identify enriched GO terms in a threshold-free
275 way, by finding GO-terms that are overrepresented at the top of a ranked list of genes. Here we ranked
276 genes by FDR for the exposure effect. GO terms were considered to be significantly enriched when p
277 < 0.05.

278

279 **Results**

280 Median coverage across the whole experiment (N = 51 BS-seq libraries) was 898X, but coverage was
281 heterogeneous between loci (Fig 1) and samples (Table S1). Mean read coverage per sample ranged
282 from 2.6X to 48.2X. Four samples had less than 10X mean coverage per CpG, however, these samples
283 were distributed across treatment groups.

284 Treatment with 5-aza-dC had wide-ranging effects on DNA methylation across the genome.
285 First, samples clustered strongly by exposure regime at the gene-level in both the t-SNE visualisation
286 (Fig 2) and in the principle component analyses (Fig 3), indicating that 5-aza-dC has a strong effect on
287 the methylation of genes. Both analyses show the strong clustering at the gene loci, and that wasps
288 exposed for 24 hours are more like controls than wasps treated for 48 hours. In addition, the PCA
289 analysis showed clear separation along PCA1 (which presumably is heavily associated with exposure
290 regime; Fig 3). Importantly though, this analysis also emphasises the extent of variation in response
291 across the three treatment regimes, which is smallest for the controls, and largest for the 48 hour-
292 exposed wasps (note the spread along PCA axis 2 in Fig 3). In addition, both cluster analyses placed
293 the controls as some way intermediate between the two 5-aza-dC exposed treatments, which we
294 consider further in the next two paragraphs. Both analyses also show that exposure regime dominated
295 the clustering, over and above of any effects of either tissue-type or collection time, suggesting that
296 this is the key effect. When we consider clustering at the level of individual CpG loci, clustering is much
297 less obvious, using both t-SNE and PCA. However, the control treatment appears to cluster within the
298 others in the t-SNE analyses, and the 24-hour and 48-hour treatments do separate to some extent,

299 e.g. with the 24h samples more the bottom and left of Fig S1 and the 48h samples more to the top
300 and right (see Figs S1 and S2). The differences between the individual CpG and gene-level cluster
301 analyses are likely to be due to a combination of low coverage and a greater stochasticity of
302 methylation at the individual CpG level (see discussion).

303 Second, a more in-depth examination of the data using a generalised linear modelling (GLM)
304 approach to examine differential methylation at the gene level found that 5-aza-dC exposure had a
305 significant effect on the methylation status of the majority of genes examined; exposure treatment
306 was significant as a main effect at an FDR level of 0.05 for 8,725 out of 11,582 genes, or 75.3% of genes
307 (see also Fig S3). Moreover, both tissue-type and collection time also influenced methylation status
308 across a large number of genes (significant as main effects in the GLMs for 3833 and 3566 genes
309 respectively). Significant interactions between factors were also present for the majority of genes
310 (exposure regime * collection time = 10,776 genes, exposure regime * tissue type = 10,484 genes,
311 collection time * tissue type = 5,364 genes, and exposure * collection time * tissue type = 10,484
312 genes), showing widespread evidence for significant interplay between these factors. These results
313 emphasise the context-dependent nature of the action of 5-aza-dC. Full results from the GLMs are
314 presented in Table S2.

315 Interestingly, it appears that methylation tends to increase initially in response to exposure to
316 5-aza-dC, an effect that is noticeable after 24h of exposure, before decreasing after 48h of exposure
317 (Fig 4A). This means that how long individuals are exposed to the chemical may well influence what
318 kind of methylation changes are observed; in other words, both hyper- and hypo-methylation can
319 result from 5-aza-dC treatment. In terms of when wasps were collected after exposure, 3,566 genes
320 displayed significantly altered methylation in association with collection time, with a more
321 pronounced decrease in methylation in samples collected 48h post-exposure relative to 24h post-
322 exposure (when compared to 0h post-exposure; Fig 4B). This suggests that the effect of 5-aza-dC
323 persists at least 48h after exposure has stopped. Finally, 3,833 genes displayed significantly altered
324 methylation patterns in association with tissue type (Fig 4C), with lower methylation overall in the
325 head. However, we re-iterate that these main effects need to be contextualised within the pattern of
326 widespread interactions seen in the data, with 91% of genes exhibiting a significant second-order
327 interaction between exposure, tissue type and collection time.

328 Gene-set enrichment analysis, using *N. vitripennis* GO terms, when genes were ranked by their
329 response to exposure regime is presented in Table S3. Broadly, we identified 169 significantly enriched
330 GO terms ($p < 0.05$), and 80 of these terms fell under the "Biological process" (BP) category, 18 under
331 "Cellular component" (CC), and 71 under "Molecular function" (MF). The top five enriched GO terms

332 were “sequence-specific DNA binding” (GO:0043565), “regulation of transcription, DNA templated”
333 (GO:0006355), “ATP binding” (GO:0005524), “DNA binding transcription factor activity” (GO:
334 0003700), and “protein kinase activity” (GO:0004672) (Table S3). When the analysis was repeated
335 using GO terms for the *D. melanogaster* orthologs, a higher number of significantly enriched GO terms
336 was obtained (700; 538 BP, CC 71, MF 91; Table S4). The top five GO terms for the *D. melanogaster*
337 analysis were “nucleus” (GO:0005634), “positive regulation of transcription by RNA polymerase II”
338 (GO:0045944), “regulation of Notch signalling pathway” (GO:0008593), “imaginal disc-derived wing
339 morphogenesis” (GO:0007476), and, making a reappearance, “sequence-specific DNA binding”
340 (GO:0043565). This analysis combined GO-terms when *N. vitripennis* genes had multiple *D.*
341 *melanogaster* orthologs, which could lead to bias, however, we retrieve very similar enriched terms
342 when we used only the GO terms that were shared between the multiple orthologs (Table S5),
343 suggesting any bias caused by this approach is small.

344

345 **Discussion**

346 A genome-wide analysis of CpGs across more than 11,000 genes, using 51 biological replicates across
347 our treatment combinations, has shown that the demethylating agent 5-aza-dC has very widespread
348 effects on patterns of DNA methylation in the *Nasonia vitripennis* genome. The methylome has been
349 characterised in *Nasonia vitripennis* by Wang *et al.* [23] and Beeler *et al.* [57], and in common with
350 other insect species, DNA methylation appears to be primarily associated with gene bodies (including
351 intron-exon boundaries), and a rather low proportion of CpGs are methylated. As mentioned above,
352 previous work that has attempted to confirm the effects of 5-aza-dC on DNA methylation have had
353 mixed results. Pegoraro *et al.* [24] confirmed the effects of 5-aza-dC across five genes (four via a qPCR
354 MethylQuant assay and one via cloning a fragment of the *msn* gene from treated versus untreated
355 individuals), whilst Ellers *et al.* [38] have recently failed to show any effects of 5-aza-dC on the
356 methylation status of 155 CpGs across ten genes (nine of which provided libraries; they used high-
357 throughput bisulfite amplicon sequencing: BSAS; see [58] for a discussion of different methylation
358 quantification assays). Our larger-scale study has both confirmed the action of 5-aza-dC as a disruptor
359 of methylation in *Nasonia*, but also suggested why it might be so difficult to pin down those effects.
360 In particular, we find that whilst we see statistically significant main effects of exposure regime on
361 DNA methylation (which includes differences in length of exposure), we also see effects of tissue type
362 (albeit at the crude level of heads versus bodies), and time since exposure to 5-aza-dC. Moreover, we
363 also see many statistically significant interactions between these factors; all told, the action of 5-aza-
364 dC is likely to be context-dependent, both in terms of the sequences more likely to be targeted but
365 also in terms of the experimental details, including target tissues. Nonetheless, our results strongly

366 suggest that 5-aza-dC is not chemically inert in terms of influencing DNA methylation in *Nasonia* –
367 indeed, quite the opposite.

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

376 A further important effect we see is that 5-aza-dC exposure can result in hyper-methylation
377 as well as hypo-methylation. This means that the characterisation of 5-aza-dC as a strictly
378 demethylating agent may be somewhat misleading. The complexity of outcomes of 5-aza-dC in
379 mammals, in particular in terms of trying to understand its action for cancer treatments, is becoming
380 increasingly well-appreciated [26]. As such, our data confirm in an insect what is being seen in
381 mammalian cells. We are not well placed to speculate on the mechanism(s) by which 5-aza-dC may
382 lead to hyper-methylation (see instead [26]), even though its hypomethylating effect is well
383 characterised chemically (see Introduction). However, if 5-aza-dC disrupts the regulation of the DNA
384 methylation machinery itself, for instance in terms of how the Dnmt genes are expressed and act, then
385 the downstream outcomes could be both hyper- and hypo-methylation. Perhaps related to this, in
386 their study of DNA methylation and the control of the diapause response in *Nasonia*, Pegoraro *et al.*
387 [24] used both 5-aza-dC and RNA interference (RNAi) to manipulate methylation, in the latter case
388 disrupting DNA methylation by knocking down both Dnmt1a and Dnmt3. Whilst both exposure to 5-
389 aza-dC and the RNAi constructs affected how diapause was influenced by exposure to short or long
390 day length, they did so in different ways. Whilst both manipulations abolished the day-length
391 response, the RNAi of Dnmt1a and Dnmt3 resulted in a change in how wasps responded to long-day
392 length (increasing the diapause response), whilst 5-aza-dC led to a change in response to short-day
393 length (decreasing the diapause response). As such, different manipulations of DNA methylation may
394 lead to different phenotypic outcomes, depending on the regulatory causes and consequences of the
395 methylation pattern itself. More generally then, it is clear that there is much that remains to be
396 elucidated in how 5-aza-dC shapes methylation patterns, in both insects and mammals [26].

397 On a more technical note, perhaps unsurprisingly we found more significant genes in the GO
398 enrichment analysis using the GO terms for *Drosophila melanogaster* orthologs. This is perhaps

399 because of the more detailed gene ontology information available for the *D. melanogaster* genome.
400 Nonetheless, the two analyses provide a similar story in terms of GO enrichment. In addition, the
401 majority of our analyses were undertaken at the level of the gene. Perhaps given the variation in
402 coverage at the level of individual CpGs, the clustering analyses were not by any means as clear cut as
403 the gene-level results. However, in the t-SNE analysis (Fig S1), the control replicates cluster together,
404 albeit within the broader distribution of the other two exposure treatment groups. The effect of
405 genomic and sequence context may be particularly relevant here: our analysis at the level of the gene
406 (see Methods for details) brings together CpGs that share a lot of context, by virtue of being associated
407 with a given genetic locus. The differences in the clustering at the individual CpG versus gene levels
408 may therefore represent the fact that the action of 5-aza-dC does appear to be non-random, and that
409 we gain clarity when we consider closely associated CpGs whose patterns of methylation, and
410 responses to 5-aza-dC, are more similar than randomly chosen CpGs across the genome.

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

417

418 **Acknowledgments**

419 We are very grateful to Mark Lammers and Jacintha Ellers for their collegiate and constructive
420 engagement with our previous DNA methylation work. We are also grateful to Matthew Arno and the
421 Edinburgh Genomics team for their sequencing and downstream support.

422

423

References

1. Ashbrook DG, Mulligan MK, Williams RW. Post-genomic behavioural genetics: From revolution to routine. *Genes Brain Behav.* 2018; 17: e12441.

2. Biémont C. From genotype to phenotype. What do epigenetics and epigenomics tell us? *Heredity.* 2010; 105: 1-3.

3. Lyko F. The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nat Rev Genet.* 2018; 19: 81-92.

4. Feng S, Cokus SJ, Zhang X, Chen P, Bostick M *et al.* Conservation and divergence of methylation patterning in plants and animals. *Proc Natl Acad Sci USA. 2010;* 107: 8689-8694.

5. Zemach A, McDaniel IE, Silva P, Zilberman D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science.* 2010; 328: 916-919.

6. Bewick AJ, Vogel KJ, Moore AJ, Schmitz RJ. Evolution of DNA methylation across Insects. *Mol Biol Evol.* 2017; 34: 654-665.

7. Glastad KM, Hunt BG, Yi SV, Goodisman MAD. DNA methylation in insects: on the brink of the epigenomic era. *Insect Mol Biol.* 2011; 20: 553-65.

8. Amarasinghe HE, Clayton CI, Mallon EB. Methylation and worker reproduction in the bumble-bee (*Bombus terrestris*). *Proc R Soc Lond B Biol Sci.* 2014; 281: 20132502.

9. Kucharski R, Maleszka J, Foret S, Maleszka R. Nutritional control of reproductive status in honeybees via DNA methylation. *Science. 2008;* 319: 1827-1830.

10. Lyko F, Foret S, Kucharski R, Wolf S, Falckenhayn C *et al.* The honey bee epigenomes: differential metjhylation of brain DNA in queens and workers. *PLoS Biol.* 2010; 8: e1000506.

11. Li Y, Zhang L, Yi Y, Hu W, Guo Z *et al.* Genome-wide DNA methylation changes associated with olfactory learning and memory in *Apis mellifera*. *Sci Rep.* 2017; 7: 17017.

12. Smith CR, Mutti NS, Jasper WC, Naidu A, Smith CD *et al.* Patterns of DNA methylation in development, division of labor and hybridisation in an ant with genetic caste determination. *PLoS One* 2012; 7: e42433.

13. Libbrecht R, Oxley PR, Keller L, Kronauer DJ. Robust DNA methylation in the clonal raider ant brain. *Curr Biol.* 2016; 26: 391-395.

14. Bonduriansky R, Day T. Nongenetic inheritance and its evolutionary implications. *Annu Rev Ecol Evol Syst.* 2009; 40: 103-125.

15. Bonduriansky R, Day T. *Extended Heredity*. Princeton: Princeton University Press; 2018.

16. Lind MI, Spagopoulou F. Evolutionary consequences of epigenetic inheritance. *Heredity.* 2018; 121: 205-209.

17. Burggren W. Epigenetic Inheritance and its role in evolutionary biology: re-evaluation and new perspectives. *Biology. 2016;* 5: 24.

18. Patten MM, Ross L, Curley JP, Queller DC, Bonduriansky R *et al.* The evolution of genomic imprinting: theories, predictions and empirical tests. *Heredity.* 2014; 113: 119-128.

19. Haig D. *Genomic imprinting and kinship.* New Brunswick: Rutgers University Press; 2002.

20. Haig D. Parental antagonism, relatedness asymmetries and genomic imprinting. *Proc R Soc Lond B Biol Sci.* 1997; 264: 1657-1662.

21. Moore T, Haig D. Genomic imprinting in mammalian development: a parental tug-of-war. Trends Genet. 1991; 7: 45-49.

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

23. Wang X, Wheeler D, Avery A, Rago A, Choi JH. Function and evolution of DNA methylation in *Nasonia vitripennis*. *PLoS Genet*. 2013; 9: e1003872.

24. Pegoraro M, Bafna A, Davies NJ, Shuker DM, Tauber E. DNA methylation changes induced by long and short photoperiods in Nasonia. *Genome Res.* 2016; 26: 203-210.

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

26. Seelan RS, Mukhopadhyay P, Pisano MM, Greene RM. Effects of 5-Aza-2'-deoxycytidine (decitabine) on gene expression. *Drug Metab Rev.* 2018; 50: 193-207.

27. Bryan J, Kantarjian H, Garcia-Manero G, Jabbour E. Pharmacokinetic evaluation of decitabine for the treatment of leukemia. *Expert Opin Drug Metab Toxicol.* 2011; 7: 661-672.

28. Momparler RL. A perspective on the comparative antileukemic activity of 5-aza-2'-deoxycytidine (Decitabine) and 5-azacytidine (Vidaza). *Pharmaceuticals* (Basel). 2012; 5: 875-881.

29. Cook N, Pannebakker BA, Tauber E, Shuker DM. DNA methylation and sex allocation in the parasitoid wasp Nasonia vitripennis. *Am Nat. 2015;* 186: 513-518.

30. Hamilton WD. Extraordinary sex ratios. *Science.* 1967*;* 156: 477-488.

31. Taylor PD, Bulmer MG. Local Mate Competition and the sex-ratio. *J Theoret Biol.* 1980; 86: 409-419.

32. Werren JH. Sex ratio adaptations to local mate competition in a parasitic wasp. *Science.* 1980; 208: 1157-1159.

33. Werren JH. Sex ratio evolution under local mate competition in a parasitic wasp. *Evolution.* 1983; 37: 116-124.

523 *34.* Shuker DM, West SA. Information constraints and the precision of adaptation: sex ratio
524 manipulation in wasps. *Proc Natl Acad Sci USA.* 2004*;* 101: 10363-10367.
525
526 35. Burton-Chellew MN, Koevoets T, Grillenberger BK, Sykes EM, Underwood SL *et al.* Facultative
527 sex ratio adjustment in natural populations of wasps: cues of local mate competition. *Am Nat.*
528 2008; 172: 393-404.
529
530 36. West SA. Sex allocation. Princeton: Princeton University Press; 2009.
531
532 37. Wild G, West SA. Genomic imprinting and sex allocation. *Am Nat.* 2009; 173: E1-14.
533
534 38. Ellers J, Visser M, Mariën J, Kraaijeveld K, Lammers M. The importance of validating the
535 demethylating effect of 5-aza-dC in model species. *Am Nat*. 2019; Forthcoming.
536
537 39. Ramos M, Wijetunga NA, McLellan AS, Suzuki M, Greally JM. DNA demethylation by 5-aza-2'-
538 deoxycytidine is imprinted, targeted to euchromatin and has limited transcriptional
539 consequences. *Epigenetics Chromatin.* 2015; 8:11.
540
541 40. Mossman D, Kim K, Scott RJ. Demethylation by 5-aza-2'-deoxycytidine in colorectal cancer
542 cells targets genomic DNA whilst promoter CpG island methylation persists. *BMC Cancer* 2010;
543 10: 366.
544
545 41. Hagemann S, Heil O, Lyko F, Brueckner B. Azacytidine and decitabine induce gene-specific and
546 non-random DNA demethylation in human cancer cell lines. *PLoS One.* 2011; 6: e17388.
547
548 *42.* Whiting AR. The biology of the parasitic wasp *Mormoniella vitripennis* [=*Nasonia brevicornis*]
549 (Walker). *Q Rev Biol.* 1967; 42: 333-406.
550
551 43. Pannebakker BA, Halligan DL, Reynolds KT, Ballantyne GA, Shuker DM *et al.* Effects of
552 spontaneous mutation accumulation on sex ratio traits. *Evolution*. 2008; 62: 1921-1935.
553
554 44. Pannebakker BA, Watt R, Knott SA, West SA, Shuker DM. The quantitative genetic basis of sex
555 ratio variation in *Nasonia vitripennis*: a QTL study. *J Evol Biol. 2011;* 24: 12-22.
556
557 45. Cook N, Trivedi U, Pannebakker BA, Blaxter M, Ritchie MG *et al.* Oviposition but not sex
558 allocation is associated with transcriptomic changes in females of the parasitoid wasp *Nasonia*
559 *vitripennis*. *G3 (Bethesda).* 2015; 5: 2885-2892.
560
561 46. Cook N, Boulton RA, Green J, Trivedi U, Tauber E *et al.* Differential gene expression is not
562 required for facultative sex allocation: a transcriptome analysis of brain tissue in the parasitoid
563 wasp *Nasonia vitripennis*. *Roy Soc Open Sci*. 2018; 5: 171718.
564
565 47. Piskala, Sorm F. Nucleic acid components and their analogues. LI. Synthesis of 1-glycosyl
566 derivatives of 5-azauracil and 5-azacytosine. *Collect Czechoslov Chem Commun.* 1964; 29:
567 2060-2076.
568
569 48. Jones PA, Taylor SM. Cellular differentiation, cytidine analogs and DNA methylation. *Cell.*
570 *1980;* 20: 85-93.
571

572 49. Momparler RL. Molecular, cellular and animal pharmacology of 5-aza-2'-deoxycytidine.
573 *Pharmacol Ther.* 1985; 30: 287-299.
574
575 50. Christman JK 5-azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation:
576 mechanistic studies and their implications for cancer therapy. *Oncogene.* 2002; 21: 5483-
577 5495.
578
579 51. Cook N, Parker DJ, Tauber E, Pannebakker BA, Shuker DM. Validating the demethylating
580 effects of 5-aza-2'-deoxycytidine in insects requires a whole-genome approach (A reply to
581 Ellers *et al.*) *Am Nat.* 2019; Forthcoming.
582
583 52. R Core Team. R: A language and environment for statistical computing. R Foundation for
584 Statistical Computing, Vienna, Austria. 2017. Available from https://www.R-project.org
585
586 53. van der Maaten LJP, Hinton GE. Visualizing high-dimensional data using t-SNE. *J Mach Learn*
587 *Res.* 2008; 9: 2579-2605.
588
589 54. Krijthe JH. Rtsne: T-Distributed Stochastic Neighbor Embedding using a Barnes-Hut
590 Implementation. 2015. Available from: https://github.com/jkrijthe/Rtsne
591
592 55. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful
593 approach to multiple testing. *J R Stat Soc Series B Stat Methodol.* 1995; 57: 289–300.
594
595 56. Rahnenfuhrer AAJ. *topGO: Enrichment Analysis for Gene Ontology*. R package version 2.28.0.
596 2016. Available from http://bioconductor.org/packages/release/bioc/html/topGO.html
597
598 57. Beeler SM, Wong GT, Zheng JM, Bush EC, Remnant EJ et al. Whole-genome DNA methylation
599 profile of the Jewel Wasp (*Nasonia vitripennis*). G3 (Bethesda). 2014; 4: 383-388.
600
601 58. Redshaw N, Huggett JF, Taylor MS, Carole A, Foy CA *et al.* Quantification of epigenetic
602 biomarkers: an evaluation of established and emerging methods for DNA methylation
603 analysis. *BMC Genomics*. 2014; 15: 1174.

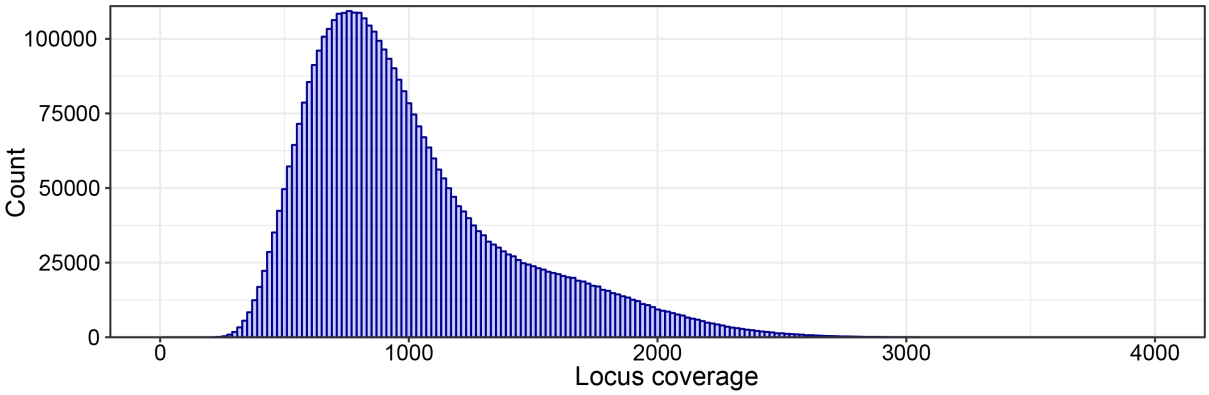
604

605 **Data and code availability**

606 Raw sequence reads and processed data (proportion methylated reads per CpG) have been deposited
607 in the GEO hosted by NCBI (Accession: GSE125388). Scripts for the analyses in this paper are available
608 at: https://github.com/DarrenJParker/5-Aza-dC_methylation.

609

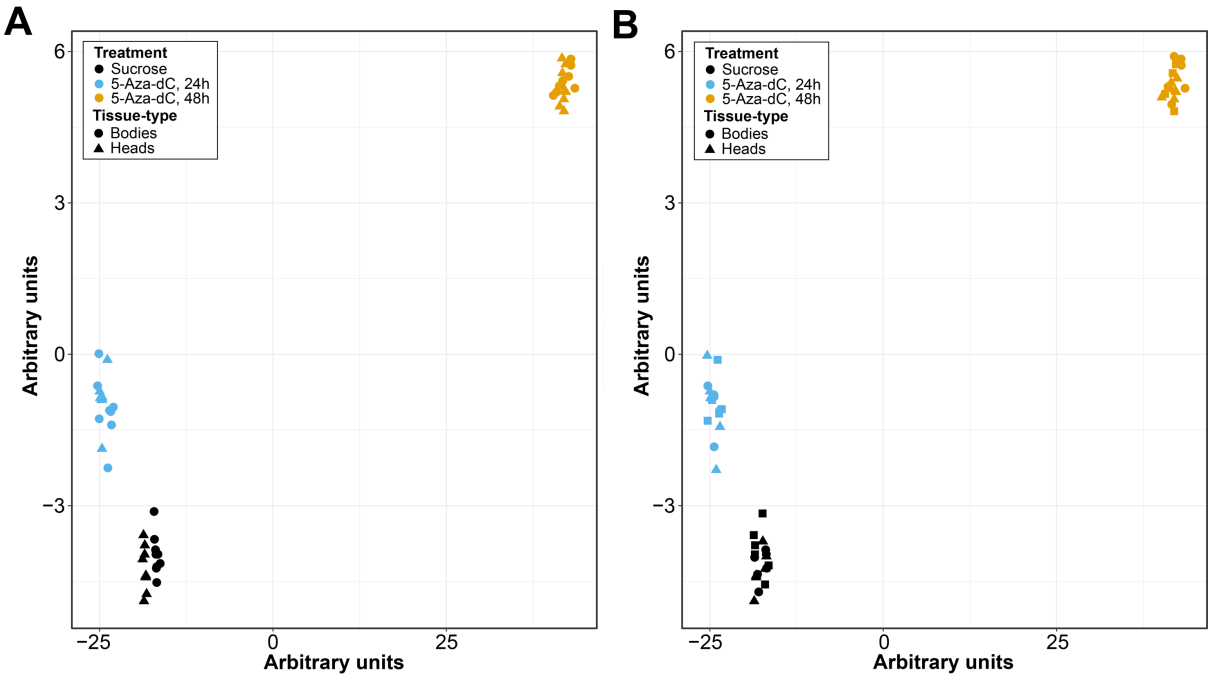
610 **Figures**



611

612 **Fig 1. Read coverage by locus across all 51 BS-Seq libraries.**

613



614

615 **Fig 2. t-SNE of the proportion of methylated reads per gene.** Shows the effects of (A) Exposure

616 regime and tissue type and (B) Exposure regime and harvest time.

617

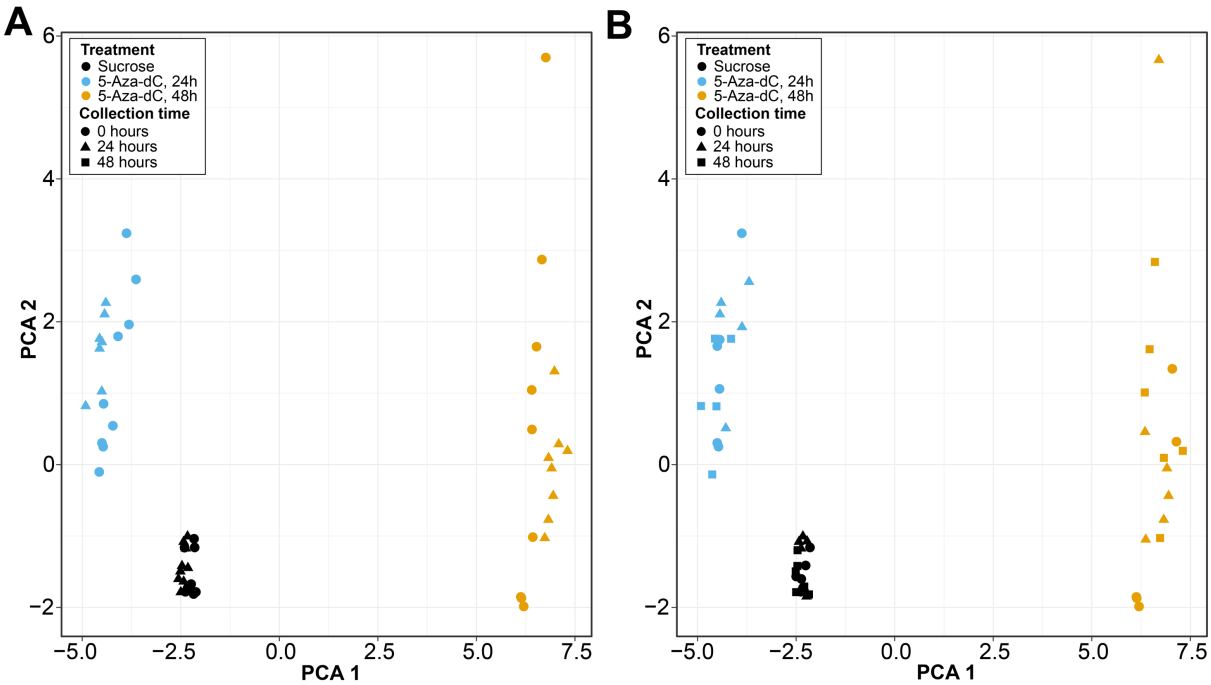


Fig 3. PCA of the proportion of methylated reads per gene. Shows the effects of (A) Exposure regime and tissue type and (B) Exposure regime and harvest time.

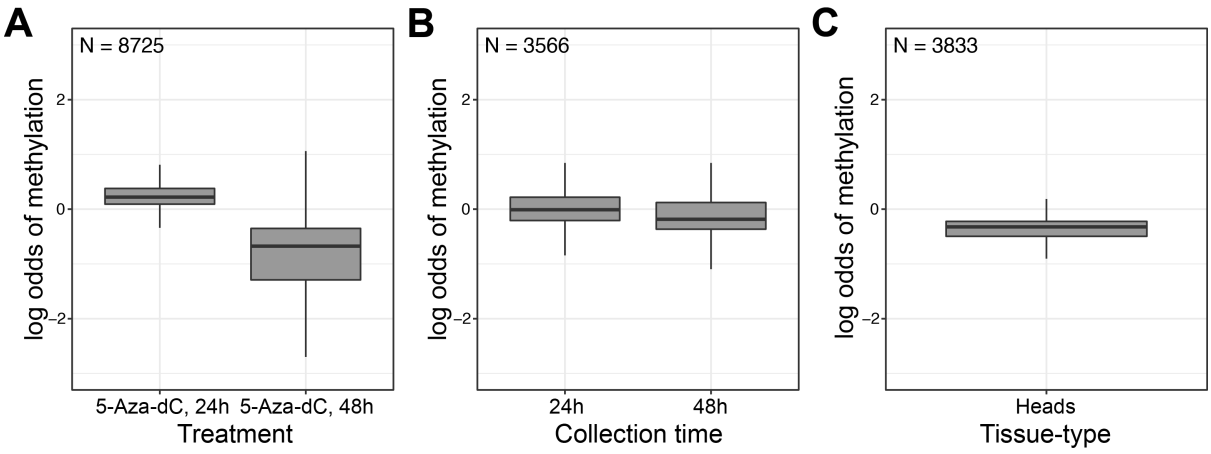


Fig 4. Gene methylation status displayed as log odds of methylation for each main effect in the experiment. A significant effect on gene methylation status was noted in response to (A) 5-aza-dC exposure regime (relative to sucrose control), (B) harvest time after exposure (relative to zero hours) and (C) tissue-type (heads relative to bodies). N= number of genes with significantly differentially methylated CpGs in association with each main effect.