

1   **Title.** Gene-regulatory independent functions for insect DNA methylation  
2

3   **Authors.** Adam J. Bewick<sup>1</sup>, Zachary Sanchez<sup>2</sup>, Elizabeth C. Mckinney<sup>2</sup>, Allen J. Moore<sup>2</sup>,  
4   Patricia J. Moore<sup>2\*</sup>, Robert J. Schmitz<sup>1\*</sup>

5   **Affiliations.** <sup>1</sup>Department of Genetics, University of Georgia, Athens, GA 30602, USA

6   <sup>2</sup>Department of Entomology, University of Georgia, Athens, GA 30602, USA

7   \*Corresponding authors. E-mail: \*[pjmoore@uga.edu](mailto:pjmoore@uga.edu) (PJM); \*[schmitz@uga.edu](mailto:schmitz@uga.edu) (RJS)

8  
9   **Abstract.** The function of cytosine (DNA) methylation in insects remains unknown.  
10 Using RNA interference we provide evidence for the functional role of the maintenance  
11 DNA methyltransferase 1 (*Dnmt1*) in *Oncopeltus fasciatus*, a hemimetabolous insect.  
12 Individuals depleted for *Dnmt1*, and subsequently DNA methylation, failed to reproduce.  
13 Depletion of DNA methylation did not result in changes in gene or transposable element  
14 expression. Eggs were inviable and declined in number, and nuclei structure of follicular  
15 epithelium was aberrant, revealing an important function of DNA methylation seemingly  
16 not contingent on gene expression. Our work provides direct experimental evidence for a  
17 functional role of *Dnmt1* and DNA methylation independent of gene expression in  
18 insects.  
19

20  
21   **Main.** Cytosine (DNA) methylation in insects has been hypothesized to play numerous  
22 functional roles including polyphenism, diapause, longevity, and social behavior and  
23 caste differentiation<sup>1–21</sup>. In many insects DNA methylation is localized to moderately and  
24 constitutively expressed genes that are highly conserved between species. This has led  
25 some to the hypothesis that DNA methylation functions in transcriptional regulation<sup>6, 15,</sup>  
26 <sup>22, 23</sup>. Evidence supporting hypotheses for the function of DNA methylation is limited due  
27 to a lack of mutant studies<sup>24</sup>. Here we show that DNA methylation is essential and is not  
28 associated with transcription in the hemipteran *Oncopeltus fasciatus* (the large milkweed  
29 bug)<sup>25, 26</sup>. Post-transcriptional knockdown of the maintenance DNA methyltransferase  
30 *Dnmt1* led to reduced egg viability, fecundity, and aberrant follicular epithelium, and thus  
31 failure to produce a successive generation. Despite finding levels of methylated CG  
32 (mCG) within coding regions reduced by 83.55%, we found no evidence for DNA  
33 methylation directly affecting transcription. Our results suggest *Dnmt1* plays an  
34 important role in reproduction in *O. fasciatus* that is mediated by a gene-regulatory  
35 independent function of DNA methylation. *Oncopeltus fasciatus* represents a fruitful  
36 model species for functional studies of DNA methylation, and continuation of studies in  
37 this system will unravel the insect epigenome and its functional consequences.

38   To assess the function of *Dnmt1*, double-stranded RNA (dsRNA) targeting *Dnmt1*  
39 (ds-*dnm1*) (Extended Data Fig. 1, and Extended Data Table 1 and 2) was injected  
40 between the abdominal sternites of virgin *O. fasciatus* females. Maternal dsRNA  
41 injection has been used in *O. fasciatus* to post-transcriptionally knockdown expression in  
42 embryos developing from eggs laid by injected females<sup>27</sup> and is known to reduce  
43 expression of embryonically expressed genes. A control injection of double-stranded *Red*  
44 (ds-*red*) or buffer was used to confirm that there was no technical affect associated with  
45 RNAi treatment. The effects of RNAi treatment on gene expression was assessed in gut,  
46 head, thorax and ovary tissues 10 days after mating to untreated *O. fasciatus* males. The

47 post-transcriptional knockdown reduced *dnmt1* mRNA expression in all examined tissues  
48 (Fig. 1a, Extended Data Fig. 2, and Extended Data Table 2). Moreover, there was little  
49 variation in expression between biological replicates in ds-*dnmt1*-injected and control  
50 individuals (Fig. 1a). Taken together, treatment with ds-*dnmt1* specifically and reliably  
51 knocked down *Dnmt1* transcripts.

52 Post-transcriptional knockdown of *Dnmt1* affects egg development and viability.  
53 In the early stages following injection, females injected with ds-*dnmt1* did not lay fewer  
54 eggs ( $F = 2.91$ ,  $df = 1$ ;  $P$  value = 1.1e-01) (Extended Data Fig. 3). Although eggs laid by  
55 females injected with ds-*dnmt1* within the first 8 days post-injection looked typical, they  
56 were significantly less likely to develop than the eggs laid by control females ( $\chi^2 = 8.470$ ;  
57  $df = 1$ ;  $P$  value = 4.0e-03). Furthermore, a mean of 93% of the eggs laid by control  
58 females initiated development whereas only a mean of 4% of the eggs laid by ds-*dnmt1*  
59 females initiated development (Fig. 1b). Although eggs laid by the control females that  
60 initiated development were viable and hatched, the few eggs laid by ds-*dnmt1*-injected  
61 females that initiated development were not viable and failed to hatch.

62 Post-transcriptional knockdown of *Dnmt1* affects egg production following the  
63 first 10-day post-injection. By 10 to 12 days, females injected with ds-*dnmt1* have mainly  
64 stopped laying eggs. While ovaries dissected from control females at this stage have  
65 intact eggs in the oviduct, the oviducts of ds-*dnmt1*-injected females are either empty or  
66 filled with a mass of what appears to be yolk (Fig. 1c), indicating a fault in production of  
67 a functional chorion. Analysis of ovarian cell structure indicates that knockdown of  
68 *Dnmt1* transcripts affects the cell structure of the follicular epithelium, which is  
69 responsible for production of the chorion and vitelline envelope. Nuclei of the follicular  
70 epithelium in *Dnmt1* post-transcriptional knock downed females are aberrant and fewer  
71 in number than that of the control females (Fig. 1c). Therefore, both through disruption of  
72 embryonic development from eggs produced within a week of injection and cessation of  
73 egg production, reproduction is compromised following knockdown of *Dnmt1*  
74 transcripts, preventing a successive generation.

75 The successful knockdown of *Dnmt1* transcripts in gut, head, thorax and ovaries  
76 prompted the evaluation of the consequences on DNA methylation using whole genome  
77 bisulfite sequencing (WGBS). We used a low coverage sequencing approach, which is  
78 sufficient for the detection of changes in bulk levels of DNA methylation<sup>28</sup>. Although  
79 *Dnmt1* mRNA expression was reduced in all tissues, DNA methylation was only reduced  
80 in ovary tissue (Fig. 2a–c, and Extended Data Fig. 2). The reduction specifically in  
81 ovaries is likely due to the higher rate of cell division in comparison to cells in other  
82 tissues surveyed, which would facilitate the passive of loss of DNA methylation in the  
83 absence of *Dnmt1*.

84 High coverage single-base resolution DNA methylomes from ds-*dnmt1* and  
85 control ovaries were generated to understand the impact of the loss of *Dnmt1*. Greater  
86 than 75% reductions of mCG were observed across the genome, consistent with the  
87 decrease observed in the low coverage experiments (Extended Data Fig. 2). The presence  
88 of symmetrical mCG – DNA methylation occurring on both DNA strands at a CpG site –  
89 is indicative of the presence of a functional *Dnmt1*. In control individuals methylation at  
90 CpGs is highly symmetrical and highly methylated. However, even though knockdown of  
91 *Dnmt1* significantly reduces methylation in ovaries, a minority of CpG sites remain  
92 symmetrically methylated (Fig. 2c). This suggests that there are some cells within ovaries

93 that have wild-type methylomes. The lack of a complete loss of mCG is expected due to  
94 the injection of fully developed individuals and hence the presence of fully methylated  
95 genomes that existed prior to the injection of dsRNA.

96 The *O. fasciatus* methylome revealed much higher levels of mCG throughout the genome,  
97 similar to other hemimetabolous insects, when compared to holometabolous  
98 insects (Fig. 2a and b). DNA methylation of gene bodies is found in *O. fasciatus* similar  
99 to other insects that possess DNA methylation, however, the pattern of mCG within  
100 exonic regions of hemi- and holometabolous insects is distinct (Fig. 2b). Higher levels of  
101 mCG are towards the 5' end compared to the 3' end of coding regions in holometabolous  
102 insects. This distribution resembles the most recent common ancestor of all insects  
103 (crustaceans) represented by *Daphnia pulex*. Hemimetabolous insects, which include *O.*  
104 *fasciatus*, have a more uniform distribution of mCG across coding regions, and higher  
105 levels towards the 3' end of coding regions. Previous studies have demonstrated that CG-  
106 methylated genes are often conserved within insects. We identified 39.31% (N = 7,561)  
107 CG-methylated genes in *O. fasciatus* and 85.99% of these are CG-methylated in at least  
108 one other insect species investigated (Extended Data Table 3). Therefore, even though the  
109 pattern of DNA methylation within gene bodies of hemimetabolous insects is distinct  
110 compared to holometabolous insects, the targeting of specific genes is conserved. Thus,  
111 the reductions of mCG across gene bodies in ds-*dnmt1* individuals provides an  
112 opportunity to study its potential function.

113 We performed RNA-seq analysis from ovary and other tissues in ds-*dnmt1* and  
114 control individuals to better understand the relationship between gene expression and  
115 DNA methylation. No relationship between discrete mCG levels and gene expression was  
116 observed and this was consistent in other insects that we investigated (Fig. 3a, Extended  
117 Data Fig. 4, and Supplementary Table 1–4). Furthermore, no correlation between  
118 continuous mCG levels and gene expression was observed (Fig. 3b, Extended Data Fig.  
119 4, and Supplementary Table 1–4). Interestingly, genome-wide relationships between  
120 DNA methylation and gene expression were not impacted by the knockdown of *Dnmt1*  
121 transcripts (Fig. 3b).

122 A maximum of 264 differentially expressed genes (DEG) were observed between  
123 *O. fasciatus* ds-*dnmt1* and control ovaries (Supplementary Table 1). *Dnmt1* was down-  
124 regulated in all ds-*dnmt1* samples and no detection of the *de novo* DNA  
125 methyltransferase *Dnmt3* was observed. Additionally, no DEGs are observed between *O.*  
126 *fasciatus* ds-*dnmt1* and control for gut, head and thorax (Supplementary Table 1 and 2).  
127 Of genes with differences in DNA methylation between ds-*dnmt1* and control ovaries (N  
128 = 6,590), the majority (N = 6,484; 98.39%) had no changes to gene expression (Fig. 3c  
129 and Supplementary Table 1 and 5). Furthermore, mCG was always reduced in ds-*dnmt1*  
130 compared to control ovaries in the 6,484 gene set. Despite genes being unmethylated  
131 (UM) in ds-*dnmt1* ovaries, changes to gene expression occurred in both directions for the  
132 106 differentially DNA methylated and expressed genes (Fig. 3d). Even for genes that by  
133 definition are unmethylated (N = 5,982) or similarly CG-methylated (N = 21) in ds-  
134 *dnmt1* and control, a similar proportion (N = 6,003; 97.53%) was observed to have no  
135 difference in gene expression. The lack of an association between DNA methylation and  
136 gene expression is further supported when applying more stringent thresholds to the  
137 definition of CG-methylated and unmethylated genes (Extended Data Fig. 5). Our  
138 observations support a trivial role of DNA methylation in gene expression if any at all.

139        The function of DNA methylation in *O. fasciatus* might lie within genome  
140 defense – the transcriptional regulation of repetitive DNA and transposons – as its  
141 genome is composed of a nontrivial amount of repetitive DNA and transposons (6.21%)  
142 (Supplementary Table 6). However, as in genes, reduction of mCG in ds-*dnmt1*  
143 compared to control ovaries was found across the bodies of the TEs, and were not  
144 associated with a genome-wide increase in expression (Fig. 4a–c, Extended Data Fig. 6  
145 and Extended Data Table 4). This further supports a trivial role of DNA methylation in  
146 transcriptional regulation of loci.

147        In summary, post-transcriptional knockdown of *Dnmt1* impacted the maternal  
148 somatic gonad and egg maturation. The knockdown of *Dnmt1* transcripts in *O. fasciatus*  
149 ovaries subsequently did not lead to transcriptome-wide changes in gene expression.  
150 DNA methylation might be required for proper mitosis, and the segregation of sister  
151 chromatids into their respective daughter cells, as there does appear to be aberrant nuclei  
152 structure in ds-*dnmt1* compared to control ovarian cells (Fig. 1c). This phenotype could  
153 be the result of epigenomic defects in scaffold/matrix attachment regions (S/MAR), as  
154 some correlate with origins of replication<sup>29</sup>. This phenotype could be exacerbated by  
155 multiple mitotic divisions of follicular epithelial cells during oogenesis<sup>30</sup>, and/or the  
156 presence of holocentric chromosomes<sup>31</sup>. We suggest that DNA methylation is more  
157 important for genome structure, integrity or other cellular processes than it is for somatic  
158 expression in *O. fasciatus*. Future work describing the epigenomic contributions to  
159 mitotic and meiotic cellular defects are now possible to study in this newly emerging and  
160 tractable model species, *O. fasciatus*.

161

## 162        References

163

- 164        1. Falckenhayn, C. et al. Characterization of genome methylation patterns in the desert  
165 locust *Schistocerca gregaria*. *J. Exp. Biol.* **216**, 1423–1429 (2013).
- 166        2. Cullen, D. A. et al. From molecules to management: mechanisms and consequences  
167 of locust phase polyphenism. *Adv. Insect Phys.* **53**, 167–285 (2017).
- 168        3. Pegoraro, M., Bafna, A., Davies, N. J., Shuker, D. M. & Tauber, E. DNA methylation  
169 changes induced by long and short photoperiods in *Nasonia*. *Genome Res.* **26**, 203–  
170 210 (2016).
- 171        4. Reynolds, J. A. Epigenetic influences on diapause. *Adv. Insect Phys.* **53**, 115–144  
172 (2017).
- 173        5. Kozeretska, I. A., Serga, S. V., Koliada, A. K. & Vaiserman, A. M. Epigenetic  
174 regulation of longevity in insects. *Adv. Insect Phys.* **53**, 87–114 (2017).
- 175        6. Bonasio, R. et al. Genome-wide and caste-specific DNA methylomes of the ants  
176 *Camponotus floridanus* and *Harpegnathos saltator*. *Curr. Biol.* **22**, 1755–1764  
177 (2012).
- 178        7. Foret, S. et al. DNA methylation dynamics, metabolic fluxes, gene splicing, and  
179 alternative phenotypes in honey bees. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 4968–4973  
180 (2012).
- 181        8. Lockett, G. A., Kucharski, R. & Maleszka, R. DNA methylation changes elicited by  
182 social stimuli in the brains of worker honey bees. *Genes Brain Behav.* **11**, 235–242  
183 (2012).

- 184 9. Ford, D. Honeybees and cell lines as models of DNA methylation and aging in  
185 response to diet. *Exp. Gerontol.* **48**, 614–619 (2013).
- 186 10. He, X. J. et al. Making a queen: an epigenetic analysis of the robustness of the  
187 honeybee (*Apis mellifera*) queen developmental pathway. *Mol. Ecol.* **26**, 1598–1607  
188 (2017).
- 189 11. Libbrecht, R., Oxley P. R., Keller L. & Kronauer, D. J. Robust DNA methylation in  
190 the clonal raider ant brain. *Curr. Biol.* **26**, 391–395 (2016).
- 191 12. Bewick, A. J., Vogel, K. J., Moore, A. J. & Schmitz, R. J. Evolution of DNA  
192 methylation across insects. *Mol. Biol. Evol.* **34**, 654–665 (2016).
- 193 13. Glastad, K. M. et al. Variation in DNA methylation is not consistently reflected by  
194 sociality in Hymenoptera. *Genome Biol. Evol.* **9**, 1687–1698 (2017).
- 195 14. Provataris, P., Meusemann, K., Niehuis, O., Grath, S. & Misof, M. Signatures of  
196 DNA methylation across insects suggest reduced DNA methylation levels in  
197 Holometabola. *Genome Biol. Evol.* **evy066** (2018).
- 198 15. Glastad, K. M., Gokhale, K., Liebig, J. & Goodisman, M. A. D. The caste- and sex-  
199 specific DNA methylome of the termite *Zootermopsis nevadensis*. *Sci. Rep.* **6**, 37110  
200 (2016).
- 201 16. Lyko, F. et al. The honey bee epigenomes: differential methylation of brain DNA in  
202 queens and workers. *PLoS Biol.* **8**, e1000506 (2010).
- 203 17. Xiang, H. et al. Single base-resolution methylome of the silkworm reveals a sparse  
204 epigenomic map. *Nat. Biotechnol.* **28**, 516–520 (2010).
- 205 18. Wang, X. et al. Function and evolution of DNA methylation in *Nasonia vitripennis*.  
206 *PLoS Genet.* **9**, e1003872 (2013).
- 207 19. Cunningham, C. B. et al. The genome and methylome of a beetle with complex social  
208 behavior, *Nicrophorus vespilloides* (Coleoptera: Silphidae). *Genome Biol. Evol.* **7**,  
209 3383–3396 (2015).
- 210 20. Patalano, S. et al. Molecular signatures of plastic phenotypes in two eusocial insect  
211 species with simple societies. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 13970–13975 (2015).
- 212 21. Rehan, S. M., Glastad, K. M., Lawson, S. P. & Hunt, B. G. The genome and  
213 methylome of a subsocial small carpenter bee, *Ceratina calcarata*. *Genome Biol.  
Evol.* **8**, 1401–1410 (2016).
- 214 22. Sarda, S., Zeng, J., Hunt, B. G. & Yi, S. V. The evolution of invertebrate gene body  
215 methylation. *Mol. Biol. Evol.* **29**, 1907–1916 (2012).
- 216 23. Hunt, B. G., Glastad, K. M., Yi, S. V. & Goodisman, M. A. D. Patterning and  
217 regulatory associations of DNA methylation are mirrored by histone modifications in  
218 insects. *Genome Biol. Evol.* **5**, 591–598 (2013).
- 219 24. Li-Byarlay, H. et al. RNA interference knockdown of DNA methyl-transferase 3  
220 affects gene alternative splicing in the honey bee. *Proc. Natl. Acad. Sci. U.S.A.* **110**,  
221 12750–12755 (2013).
- 222 25. Chipman, A. D. *Oncopeltus fasciatus* as an evo-devo research organism. *Genesis* **55**  
223 (2017).
- 224 26. Panfilio, K. A. et al. Molecular evolutionary trends and feeding ecology  
225 diversification in the Hemiptera, anchored by the milkweed bug genome. Preprint at  
226 <https://www.biorxiv.org/content/early/2017/10/11/201731> (2017).
- 227

- 228 27. Liu, P. Z. & Kaufman, T. C. Hunchback is required for suppression of abdominal  
229 identity, and for proper germband growth and segmentation in the intermediate  
230 germband insect *Oncopeltus fasciatus*. *Development* **131**, 1515–1527 (2004).
- 231 28. Bewick, A. J. et al. FASTmC: A suite of predictive models for nonreference-based  
232 estimations of DNA methylation. *G3 (Bethesda)* **6**, 447–752 (2015).
- 233 29. Vaughn, J. P., Dijkwel, P. A., Mullenders, L. H. & Hamlin, J. L. Replication forks are  
234 associated with the nuclear matrix. *Nucleic Acids Res.* **18**, 1965–1969 (1990).
- 235 30. Bonhag, P. F. & Wick, J. R. The functional anatomy of the male and female  
236 reproductive systems of the milkweed bug, *Oncopeltus fasciatus* (Dallas)  
237 (Heteroptera: Lygaeidae). *J. Morphol.* **93**, 177–283 (1953).
- 238 31. Drinnenberg, I. A., deYoung, D., Henikoff, S. & Malik, H. S. Recurrent loss of  
239 CenH3 is associated with independent transitions to holocentricity in insects. *Elife*  
240 doi: 10.7554/eLife.03676 (2014).

241

242 **Acknowledgements.** We thank Sam Arsenault (University of Georgia [UGA]), and  
243 Mariana Monteiro and Stefan Götz (Blast2GO Team) for assistance with Blast2GO  
244 analyses, Nick Rohr and Tina Ethridge for MethylC-Seq and RNA-seq library  
245 preparation, Brigitte Hofmeister (UGA) for bioinformatics advice, Mary Goll (UGA) for  
246 comments, Kelly Dawe (UGA) for feedback on cellular phenotypes, Muthugapatti  
247 Kandasamy of the Biomedical Microscopy Core (UGA) for technical assistance with  
248 confocal microscopy, and Cassandra Extavour (Harvard) for providing technical advice  
249 and *Red* plasmids for RNAi. We also thank the Georgia Advanced Computing Resource  
250 Center (GACRC), Georgia Genomics and Bioinformatics Core (GGBC) and the Office of  
251 Research at the University of Georgia. This work was funded by the National Science  
252 Foundation (NSF) IOS-1354358 (AJM), and UGA CAES Undergraduate Research  
253 Award (ZS). RJS is a Pew Scholar in the Biomedical Sciences, supported by The Pew  
254 Charitable Trusts.

255

256 **Authors contributions.** AJB, RJS, PJM, and AJM designed the study. ZS and PJM  
257 performed microscopy analyses. ECM performed RNAi, and extracted DNA and RNA  
258 for library preparation and sequencing. AJB analyzed the MethylC-seq and RNA-seq data,  
259 and wrote the manuscript with contributions from RJS, PJM, AJM and ECM.

260

261 **Author information.** The authors declare no competing interests.

262

263 **Materials and correspondence.** Correspondence and requests for materials should be  
264 addressed to [pjmoore@uga.edu](mailto:pjmoore@uga.edu) and [schmitz@uga.edu](mailto:schmitz@uga.edu).

265

266

267

268

269

270

271

272

273

274 **Figure legends**

275

276 **Fig. 1: *Dnmt1* is required for reproduction in *O. fasciatus*.** **a**, Assessment of RNAi  
277 treatment targeting *Dnmt1* using qRT-PCR demonstrates successful reduction in  
278 transcription in ovaries compared to control. Dots indicate mean expression level, and  
279 error bars indicate standard error of the mean. **b**, Parental RNAi injection with ds-*dnm1*  
280 significantly affected the development of eggs laid by injected females compared to eggs  
281 laid by control females. Dots indicate mean expression level, and error bars indicate  
282 standard error of the mean. **c**, Whole ovaries from females removed 12-14 days post-  
283 injection (i and ii). In control females (i), mature oocytes can be seen collecting in the  
284 lateral oviduct (Od). In ds-*dnm1* females (ii), no mature oocytes are apparent, but the  
285 oviduct has filled with a yolk-like substance. Scale bar equals 1 mm in (i) and (ii). High  
286 magnification of the follicular epithelium surrounding a maturing oocyte (iii and iv).  
287 Nuclei are stained with Hoechst 33258 and artificially colored turquoise. Nuclei from  
288 control follicular epithelium (iii) are round and regular in shape. Nuclei from ds-*dnm1*  
289 females (iv) are highly irregular in shape. Scale bar equals 20  $\mu$ m in (iii) and (iv).

290

291 **Fig. 2: *Dnmt1* is required for mCG in *O. fasciatus*.** **a**, Level and genomic location of  
292 mCG between hemi- and holometabolous insects, and *D. pulex*, and RNAi treatment  
293 targeting *Dnmt1* in *O. fasciatus*. **b**, Levels of mCG across gene bodies and one kilobase  
294 pairs (1kb) flanking sequence of hemi- and holometabolous insects, and *D. pulex*. Also  
295 shown are the levels of mCG across gene bodies for *O. fasciatus* ds-*dnm1*. **c**, Density  
296 plot representation of mCG for RNAi treated *O. fasciatus* and species investigated in this  
297 study.

298

299 **Fig. 3: Loss of mCG in *O. fasciatus* ovaries has a limited effect on transcription.** **a**,  
300 Gene expression level for deciles of increasing mCG (1–10) and unmethylated genes  
301 (UM). Error bars represent 95% confidence interval of the mean. **b**, Regression of gene  
302 expression against a continuous measure of mCG with >0 FPKM for the same set of  
303 genes that are CG-methylated in *O. fasciatus* control, but unmethylated in ds-*dnm1*. Raw  
304 P values are provided for each regression, and significance or non-significance (NS) is  
305 indicated in brackets following Bonferroni correction. **c**, Combinational overlap of genes  
306 that are differential CG-methylated and expressed, and similarly CG-methylated and  
307 expressed between *O. fasciatus* ds-*dnm1* and control. Gene groups: Differentially  
308 Methylated Gene (DMG), Differentially Expressed Gene (DEG), Similarly Methylated  
309 Gene (SMG)/UnMethylated Genes (UMG), and non-Differentially Expressed Gene (non-  
310 DEG). **d**, A heatmap showing gene expression changes for genes that are differentially  
311 CG-methylated between *O. fasciatus* ds-*dnm1* and control. Expression was standardized  
312 by the highest value per gene per biological replicate to produce a Relative Fragments Per  
313 Kilobase of transcript per Million (RFPKM) value. RFPKM were clustered using a  
314 hierarchical clustering method.

315

316 **Fig. 4: Large-scale reactivation of TEs did not follow severe reductions of mCG in**  
317 *O. fasciatus* ovaries. **a**, CG methylation levels for the top ten most abundant TEs of  $\geq 500$   
318 bp in the *O. fasciatus* genome. The dashed lines correspond to the intergenic mCG level  
319 of *O. fasciatus* ds-*dnm1* and control. **b**, Levels of mCG across the bodies and 1kb

320 flanking sequence of TEs for a single representative of DNA transposons (Chapaev),  
321 LTR retrotransposons (Gypsy), and non-LTR retrotransposons (Jockey). **c**, Expression  
322 quantified as RPKM for the top ten most abundant TEs of  $\geq 500$  bp. The subset presents  
323 the RPKM distribution when all TEs are considered within ds-dnmt1 and control ovaries.  
324

## 325 Methods

326

327 **Phylogenetic analysis.** A subset of DNA methyltransferase (*Dnmt*) 1, 2, and 3 sequences  
328 was obtained from<sup>16</sup> for phylogenetic analysis. The subset included only insect species  
329 with available MethylC-seq data, and species representatives from the dipteran suborders  
330 Brachycera and Nematocera: *Acyrtosiphon pisum*, *Aedes aegypti*, *Aedes albopictus*,  
331 *Anopheles gambiae*, *Apis mellifera*, *Bombyx mori*, *Camponotus floridanus*, *Copidosoma*  
332 *floridanum*, *Culex pipiens quinquefasciatus*, *Drosophila melanogaster*, *Harpegnathos*  
333 *saltator*, *Microplitis demolitor*, *Nasonia vitripennis*, *Nicrophorus vespilloides*, *O.*  
334 *fasciatus*, *Ooceraea (Cerapachys) biroi*, *Polistes canadensis*, *Polistes dominula*,  
335 *Solenopsis invicta*, *Tribolium castaneum*, and *Zootermopsis nevadensis*. DNA  
336 methyltransferases were reassessed in *O. fasciatus* by using InterProScan v5.23-62.0<sup>32</sup> to  
337 identify annotated proteins with a C-5 cytosine-specific DNA methylase domain  
338 (PF00145). Sequence identifiers are located in Extended Data Fig. 1a. Full-length protein  
339 sequences were aligned using PASTA v1.6.4<sup>33</sup>, and manually trimmed of divergent, non-  
340 homologous sequence in Mesquite v3.2<sup>34</sup>. Phylogenetic relationship among *Dnmt*  
341 sequences was estimated using BEAST v2.3.2<sup>35</sup> with a Blosum62+Γ model of amino acid  
342 substitution. A Markov Chain Monte Carlo (MCMC) was ran until stationarity and  
343 convergence was reached (10,000,000 iterations), and a burnin of 1,000,000 was used  
344 prior to summarizing the posterior distribution of tree topologies. A consensus tree was  
345 generated using TreeAnnotator v2.3.2, visualized in FigTree v1.4.2  
346 (<http://tree.bio.ed.ac.uk/software/figtree/>) and exported for stylization in Affinity  
347 Designer v1.5.1 (<https://affinity.serif.com/en-us/>).  
348

349

349 **PCR confirmation for the presence of a single *Dnmt1* ortholog in *O. fasciatus*.** To  
350 determine if OFAS015351 and OFAS018396 were two parts of a single *Dnmt1* ortholog,  
351 we designed one sense primer at the 3' end of OFAS015351 (Of\_DMNT1-1\_3603S;  
352 Extended Data Table 1) and two antisense primers at the 3' end of OFAS018396  
353 (Of\_DNMT1-2\_424A and Of\_DNMT1-2\_465A; Extended Data Table 1). A fourth sense  
354 primer (Of\_DNMT1-2\_1S; Extended Data Table 1) was designed at the 5' end of  
355 OFAS018396 to confirm the size and sequence of this possibly truncated gene  
356 annotation.

357

357 Polymerase Chain Reaction (PCR) with primer combinations Of\_DMNT1-  
358 1\_3603S–Of\_DNMT1-2\_424A, Of\_DMNT1-1\_3603S–Of\_DNMT1-2\_465A, and  
359 Of\_DNMT1-2\_1S–Of\_DNMT1-2\_424A was performed using Q5 Polymerase (New  
360 England BioLabs, Ipswich, MA) per manufactures instructions. Thermacycler conditions  
361 were 98°C for 15 seconds (s) (denaturing), 60°C for 30 s (annealing), and 72°C for 30 s  
362 (extension), and repeated for 40 cycles. The PCR products were then purified using  
363 QIAquick PCR Purification Kit (Qiagen, Venlo, The Netherlands), and sequenced at the  
364 Georgia Genomics and Bioinformatics Core (Athens, GA).  
365

366 **Animal culture.** *Oncopeltus fasciatus* cultures were originally purchased from Carolina  
367 Biologicals (Burlington, NC). Mass colonies were maintained in incubators under a 12  
368 h:12 h light/dark cycle at 27°C. Colonies and individual experimental animals were fed  
369 organic raw sunflower seeds and provided with *ad libitum* deionized water. Late instar  
370 nymphs were separated from the mass colonies and housed under the same conditions.  
371 Nymph colonies were checked daily for newly emerged adults. Adults were separated by  
372 sex and kept with food and water for 7–10 days until females reached sexual maturity.  
373

374 **Parental RNAi.** Template for the *in vitro* transcription of reactions was prepared from a  
375 PCR reaction in which T7 phage promoter sequences were added to the gene-specific  
376 *Dnmt1* primers<sup>36</sup>. For our control sequence, we used the red fluorescent protein (*Red*)  
377 sequence used in previous parental RNAi experiments in *O. fasciatus*<sup>36</sup> or buffer. Primer  
378 sequences can be found in Extended Data Table 2. Sense and anti-sense RNA was  
379 synthesized in a single reaction using the Ambion MEGAscript kit (ThermoFisher Sci,  
380 Waltham, MA). After purification, the double-stranded RNA (dsRNA) concentration was  
381 adjusted to 2 µg/µL in injection buffer (5 mM KCl, 0.1 mM NaH<sub>2</sub>PO<sub>4</sub>)<sup>36</sup>. Females were  
382 injected with 5 µL of dsRNA between the abdominal sternites using an insulin syringe.  
383 Following injection, females were paired with an un-injected male to stimulate  
384 oviposition and fertilize eggs. Individual females with their mate were housed in petri  
385 dishes with sunflower seeds, water and cotton wool as an oviposition site. The parental  
386 RNAi protocol has been reported to result in 100% penetrance by the third clutch of  
387 eggs<sup>36</sup> and this was also our experience.  
388

389 **Reproductive phenotype screening and analysis.** Eggs were collected between days 4–  
390 10 post-injection and assessed for development. *Oncopeltus fasciatus* embryos change  
391 from a creamy white color to orange as they develop, which indicates viability. Thus,  
392 color change is a useful tool for assessing healthy development. We examined the  
393 number of developing eggs at 5 days post-oviposition, at which point viable eggs are  
394 clearly distinguishable from inviable eggs, as well as hatching rate of eggs from females  
395 injected with double-stranded *Dnmt1* (ds-*dnm1*) and double-stranded *Red* (ds-*red*). The  
396 data for the number of eggs laid was normally distributed so differences among RNAi  
397 treated groups were tested using analysis of variance (ANOVA). The development data,  
398 however, was not normally distributed and consisted of binary states (developed and not  
399 developed), and so differences among treatment groups were analyzed with a Generalized  
400 Linear Model (GLM) using a Poisson distribution.

401 A second set of *O. fasciatus* females were injected in the same manner as  
402 described in **Parental RNAi** to assess post-transcriptional knockdown of *Dnmt1* on  
403 ovarian structure. *Oncopeltus fasciatus* females were dissected 10 days after injection. By  
404 10 days post-injection *O. fasciatus* females are beginning to stop laying recognizable  
405 eggs. Ovaries were removed from *O. fasciatus* females and placed in 1× PBS. Whole  
406 ovaries were imaged with a Leica DFC295 stereomicroscope using Leica Application  
407 Suite morphometric software (LAS V4.1; Leica, Wetzlar, Germany). Dissected ovaries  
408 were fixed within 15 minutes of dissection in 4% formaldehyde in 1× PBS for 25  
409 minutes. Fixed ovaries were stained with Hoechst 33342 (Sigma Aldrich) at 0.5 µg/ml.  
410 The stained ovarioles were imaged using a Zeiss LSM 710 Confocal Microscope (Zeiss)  
411 at the University of Georgia Biomedical Microscopy Core.

412  
413 **Quantitative RT-PCR.** To assess the effectiveness of post-transcriptional knockdown of  
414 *Dnmt1*, females were dissected 11 days post-injection. Ovaries were removed from each  
415 female, flash frozen in liquid nitrogen and stored at -80°C until processing. Total RNA  
416 (and DNA) was extracted from a single ovary per female using a Qiagen Allprep  
417 DNA/RNA Mini Kit (Qiagen, Venlo, The Netherlands) per manufacturer's instructions.  
418 Complementary DNA (cDNA) was synthesized from 500 ng RNA with qScript cDNA  
419 SuperMix (Quanta Biosciences, Gaithersburg, MD).

420 Expression level of *Dnmt1* was quantified by quantitative real-time PCR (qRT-  
421 PCR). Primers were designed for *Dnmt1* using the *O. fasciatus* genome as a reference  
422 (37). Actin and GAPDH were used as endogenous reference genes. Primer sequences can  
423 be found in Extended Data Table 2. We used Roche LightCycler 480 SYBR Green  
424 Master Mix with a Roche LightCycler 480 (Roche Applied Science, Indianapolis, IN) for  
425 qRT-PCR. All samples were run with 3 technical replicates using 10 µL reactions using  
426 the manufacturer's recommended protocol. Primer efficiency calculations, genomic  
427 contamination testing and endogenous control gene selection were performed as  
428 described by<sup>37</sup>. We used the  $\Delta\Delta CT$  method<sup>38</sup> to examine differences in expression  
429 between ds-*dnm1* and ds-*red* injected females<sup>37</sup>.

430  
431 **Whole-Genome Bisulfite Sequencing (WGBS) and analysis of cytosine (DNA)  
432 methylation.** MethylC-seq libraries for an *O. fasciatus* ds-*dnm1* and control individual  
433 were prepared according to the protocol described in<sup>39</sup> using genomic DNA extracted  
434 from ovaries (see **Materials and Methods** section **Quantitative RT-PCR**). Libraries  
435 were single-end 75 bp sequenced on an Illumina NextSeq500 machine<sup>40, 41</sup>.  
436 Unmethylated lambda phage DNA was used to as a control for sodium bisulfite  
437 conversion, and an error rate of ~0.05% was estimated. *Oncopeltus fasciatus* ds-*dnm1*  
438 and control were sequenced to a depth of ~18× and ~21×, which corresponded to an  
439 actual mapped coverage of ~9× and ~11×, respectively. Additionally, low pass (<1×)  
440 WGBS from gut, head, ovary, and thorax was performed for three control and ds-*dnm1*  
441 biological replicates. *Blattella germanica* was additionally sequenced to generate equal  
442 numbers of hemi- and holometabolous insects investigated in this study. However, DNA  
443 was extracted from whole-body minus gastrointestinal tract. MethylC-seq libraries were  
444 prepared and sequenced identically to *O. fasciatus*. An error rate of ~0.14% was  
445 estimated from unmethylated lambda phage DNA. *Blattella germanica* was sequenced to  
446 a depth of ~8×, which corresponded to an actual mapped coverage of ~5×. WGBS data  
447 for *O. fasciatus* and *B. germanica* can be found on Gene Expression Omnibus (GEO)  
448 under accession GSE109199. Previously published WGBS data for *A. mellifera*<sup>42</sup>, *Bo.  
449 mori*<sup>17</sup>, *Daphnia pulex*<sup>43</sup>, *N. vespilloides*<sup>19</sup>, and *Z. nevadensis*<sup>15</sup> were downloaded from the  
450 Short Read Archive (SRA) using accessions SRR445803–4, SRR027157–9,  
451 SRR1552830, SRR2017555, and SRR3139749, respectively. Thus, DNA methylation  
452 was investigated for six insects from six different orders spread evenly across  
453 developmental groups, and a crustacean outgroup. WGBS data was aligned to each  
454 species respective genome assembly using the methylpy pipeline<sup>44</sup>. In brief, reads were  
455 trimmed of sequencing adapters using Cutadapt v1.9<sup>45</sup>, and then mapped to both a  
456 converted forward strand (cytosines to thymines) and converted reverse strand (guanines

457 to adenines) using bowtie v1.1.1<sup>46</sup>. Reads that mapped to multiple locations, and clonal  
458 reads were removed.

459 Weighted DNA methylation was calculated for CG sites by dividing the total  
460 number of aligned methylated reads by the total number of methylated plus unmethylated  
461 reads<sup>47</sup>. For genic metaplots, the gene body (start to stop codon), 1000 base pairs (bp)  
462 upstream, and 1000 bp downstream was divided into 20 windows proportional windows  
463 based on sequence length (bp). Weighted DNA methylation was calculated for each  
464 window and then plotted in R v3.2.4 (<https://www.r-project.org/>). CG sequence context  
465 enrichment for each gene was determined through a binomial test followed by Benjamini-  
466 Hochberg false discovery rate<sup>48, 49</sup>. A background mCG level was determined from all  
467 coding sequence, which was used as a threshold in determining significance with a False  
468 Discovery Rate (FDR) correction. Genes were classified as CG-methylated if they had  
469 reads mapping to at least 20 reads mapping to 20 CG sites and a *q* value < 0.05. Using a  
470 binomial test can lead to false-negatives – highly CG-methylated genes that are classified  
471 as unmethylated (UM) – due to a low number of statistically CG-methylated sites  
472 (Supplementary Table 5). Genes classified as unmethylated, but had a mCG level greater  
473 than the lowest CG-methylated gene were dropped from future analyses.  
474

475 **Ortholog identification.** Best BLASTp hit (arguments: -max\_hsps 1 -max\_target\_seqs 1  
476 evalue 1e-03) was used to identify orthologs between *O. fasciatus* and other insect  
477 species investigated.  
478

479 **RNA-seq and differential expression analysis.** RNA-seq libraries for RNA extracted  
480 from ovaries of three biological *O. fasciatus* ds-*dnmt1* and control replicates at 11 days  
481 post-injection were constructed using Illumina TruSeq Stranded RNA LT Kit (Illumina,  
482 San Diego, CA) following the manufacturer's instructions with limited modifications.  
483 RNA from ovaries of an additional three biological *O. fasciatus* ds-*dnmt1* and control  
484 replicates, and three biological *O. fasciatus* ds-*dnmt1* and control replicates from gut,  
485 head, and thorax were extracted. The starting quantity of total RNA was adjusted to 1.3  
486 µg, and all volumes were reduced to a third of the described quantity. Libraries were  
487 single-end 75 bp sequenced on an Illumina NextSeq500 machine. RNA-seq data for *O.*  
488 *fasciatus* ds-*dnmt1* and control can be found on GEO under accession GSE109199.  
489 Previously published RNA-seq data for *A. mellifera*<sup>50</sup>, and *Z. nevadensis*<sup>15</sup> were  
490 downloaded from the SRA using accessions SRR2954345, and SRR3139740,  
491 respectively.

492 Raw RNA-seq FASTQ reads were trimmed for adapters and preprocessed to  
493 remove low-quality reads using Trimmomatic v0.33 (arguments: LEADING:10  
494 TRAILING:10 MINLEN:30)<sup>51</sup> prior to mapping to the *O. fasciatus* v1.1 reference  
495 genome assembly. Reads were mapped using TopHat v2.1.1<sup>52</sup> supplied with a reference  
496 General Features File (GFF) to the *O. fasciatus* v1.1 reference genome assembly<sup>26</sup>, and  
497 with the following arguments: -I 20000 --library-type fr-firststrand --b2-very-sensitive.

498 Differentially expressed genes (DEGs) between ds-*dnmt1* and control libraries  
499 were determined using edgeR v3.20.1<sup>53</sup> implemented in R v3.2.4 (<https://www.r->  
500 [project.org/](https://www.r-project.org/)). Genes were retained for DEG analysis if they possessed a Counts Per  
501 Million (CPM) ≥ 1 in at least ≥ 2 libraries. Significance was determined using the  
502 glmQLFTest function, which uses empirical Bayes quasi-likelihood F-tests. Parameter

503 settings were determined following best practices for DEG analysis as described by<sup>54</sup>.  
504 Gene expression metrics for *A. mellifera*, *Z. nevadensis*, and *O. fasciatus* ds-*dnmt1* and  
505 control are located in Supplementary Table 1–4, respectively.  
506

507 **Gene Ontology (GO) annotation and enrichment.** GO terms were assigned to *O.*  
508 *fasciatus* v1.1 gene set<sup>26</sup> through combining annotations from Blast2Go PRO v4.1.9<sup>55</sup>,  
509 InterProScan v5.23-62.0 (arguments: -goterms -iprlookup -appl CDD,Pfam)<sup>32</sup>, and  
510 through sequence homology to *D. melanogaster* using BLASTp (arguments: -eval 1.0e-03  
511 -max\_target\_seqs 1 -max\_hsps 1). 11,105/19,615 gene models were associated  
512 with at least one GO term and a total of 8,190 distinct GO identifiers were mapped. GO  
513 terms are found in Supplementary Table 7. Enriched GO terms in gene groups were  
514 evaluated using topGO v2.30.0<sup>56</sup> implemented in R v3.2.4 (<https://www.r-project.org/>),  
515 and significance (*P* value < 0.05) of terms was assessed using Fisher's exact test with a  
516 weighted algorithm. Gene groups were contrasted to all *O. fasciatus* genes associated  
517 with GO terms (Supplementary Table 8).  
518

519 **Transposable element (TE) annotation and expression.** Transposable elements (TEs)  
520 were identified using RepeatMasker v4.0.5 (<http://www.repeatmasker.org>) provided with  
521 the invertebrate repeat library from Repbase (<http://www.girinst.org/repbase/>)  
522 (arguments: -lib <Repbase invertebrate library> -no\_is -engine wublast -a -inv -x -gff).  
523 Following RepeatMasker, neighboring TEs of the same type were collapsed into a single  
524 locus within the outputted GFF. The unmodified GFF is located in Supplementary Table  
525 6.

526 To quantify expression from TEs RNA-seq libraries from *O. fasciatus* ds-*dnmt1*  
527 and control were independently combined and mapped to the *O. fasciatus* v1.1 reference  
528 genome assembly<sup>26</sup> using bowtie2 v2.2.9<sup>57</sup> with the following arguments: --sensitive.  
529 Mapped reads overlapping with the top ten most abundant TEs of ≥500 bp in length were  
530 identified using the *intersect* command in BEDTools suite v 2.26.0<sup>58</sup>. TE expression is  
531 quantified as Reads Per Kilobase per Million mapped reads (RPKM) for each intersected  
532 TE type by counting the number reads and dividing by the mapped library read number in  
533 millions. Significance in expression of TEs between ds-*dnmt1* and control tissues was  
534 assessed using the Mann-Whitney test with the alternative hypothesis set to “greater” in  
535 R v3.2.4 (<https://www.r-project.org/>).  
536

## 537 **References**

- 539 32. Jones, P. et al. InterProScan 5: genome-scale protein function classification.  
540 *Bioinformatics* **30**, 1236–1240 (2014).
- 541 33. Mirarab, S. et al. PASTA: ultra-large multiple sequence alignment for nucleotide and  
542 amino-acid sequences. *J. Comput. Biol.* **22**, 377–386 (2015).
- 543 34. Maddison, W. P. & Maddison, D. R. Mesquite: a modular system for evolutionary  
544 analysis. Version 3.31 <http://mesquiteproject.org> (2017).
- 545 35. Bouckaert, R. et al. BEAST 2: a software platform for Bayesian evolutionary  
546 analysis. *PLoS Comput. Biol.* **10**, e1003537 (2014).

- 547 36. Ewen-Campen, B., Jones, T. E. & Extavour, C. G. Evidence against a germplasm in  
548 the milkweed bug *Oncopeltus fasciatus*, a hemimetabolous insect. *Biol. Open* **2**, 556–  
549 568 (2013).
- 550 37. Cunningham, C. B., Douthit, M. K. & Moore, A. J. Octopaminergic gene expression  
551 and flexible social behaviour in the subsocial burying beetle *Nicrophorus*  
552 *vespilloides*. *Insect Mol. Biol.* **23**, 391–404 (2014).
- 553 38. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-  
554 time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402–408  
555 (2001).
- 556 39. Urich, M. A., Nery, J. R., Lister, R., Schmitz, R. J. & Ecker, J. R. MethylC-seq  
557 library preparation for base-resolution whole-genome bisulfite sequencing. *Nat.*  
558 *Protoc.* **10**, 475–483 (2015).
- 559 40. Cokus, S. J. et al. Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals  
560 DNA methylation patterning. *Nature* **452**, 215–219 (2008).
- 561 41. Lister, R. et al. Highly integrated single-base resolution maps of the epigenome in  
562 *Arabidopsis*. *Cell* **133**, 523–536 (2008).
- 563 42. Herb, B. R. et al. Reversible switching between epigenetic states in honeybee  
564 behavioral subcastes. *Nat. Neurosci.* **15**, 1371–1373 (2012).
- 565 43. Asselman, J., De Coninck, D. I. M., Pfrenger, M. E. & De Schamphelaere, K. A. C.  
566 Gene body methylation patterns in *Daphnia* are associated with gene family size.  
567 *Genome Biol. Evol.* **8**, 1185–1196 (2016).
- 568 44. Schultz, M. D. et al. Human body epigenome maps reveal noncanonical DNA  
569 methylation variation. *Nature* **523**, 212–216 (2015).
- 570 45. Martin, M. & Marcel, M. Cutadapt removes adapter sequences from high-throughput  
571 sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
- 572 46. Langmead, B., Trapnell, C. & Salzberg, S. L. Ultrafast and memory-efficient  
573 alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25  
574 (2009).
- 575 47. Schultz, M. D., Schmitz, R. J. & Ecker, J. R. 'Leveling' the playing field for analyses  
576 of single-base resolution DNA methylomes. *Trends Genet.* **28**, 583–585 (2012).
- 577 48. Takuno, S. & Gaut, B. S. Body-methylated genes in *Arabidopsis thaliana* are  
578 functionally important and evolve slowly. *Mol. Biol. Evol.* **1**, 219–227 (2012).
- 579 49. Niederhuth, C. E. et al. Widespread natural variation of DNA methylation within  
580 angiosperms. *Genome Biol.* **17**, 194 (2016).
- 581 50. Vleurinck, C., Raub, S., Sturgill, D., Oliver, B. & Beye, M. Linking genes and brain  
582 development of honeybee workers: a whole-transcriptome approach. *PLoS One* **11**,  
583 e0157980 (2016).
- 584 51. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina  
585 sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
- 586 52. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of  
587 insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
- 588 53. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package  
589 for differential expression analysis of digital gene expression data. *Bioinformatics* **26**,  
590 139–140 (2010).

- 591 54. Chen, Y., Lun, A. T. L. & Smyth, G. K. From reads to genes to pathways: differential  
592 expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-  
593 likelihood pipeline. *F1000Res* **5**, 1438 (2016).
- 594 55. Götz, S. et al. High-throughput functional annotation and data mining with the  
595 Blast2GO suite. *Nucleic Acids Res.* **36**, 3420–3435 (2008).
- 596 56. Alexa, A. & Rahnenfuhrer, J. topGO: Enrichment analysis for gene ontology. R  
597 package version 2.30.0 (2016).
- 598 57. Langmead, B. & Salzberg, S. Fast gapped-read alignment with Bowtie 2. *Nat  
599 Methods* **9**, 357–359 (2012).
- 600 58. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing  
601 genomic features. *Bioinformatics* **26**, 841–842 (2010).







