Title. Gene-regulatory independent functions for insect DNA methylation

Authors. Adam J. Bewick[1], Zachary Sanchez[2], Elizabeth C. Mckinney[2], Allen J. Moore[2], Patricia J. Moore[2*], Robert J. Schmitz[1*]

Affiliations. [1]Department of Genetics, University of Georgia, Athens, GA 30602, USA
[2]Department of Entomology, University of Georgia, Athens, GA 30602, USA
*Corresponding authors. E-mail: *pjmoore@uga.edu (PJM); *schmitz@uga.edu (RJS)

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

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

To assess the function of *Dnmt1*, double-stranded RNA (dsRNA) targeting *Dnmt*1 (ds-*dnmt1*) (Extended Data Fig. 1, and Extended Data Table 1 and 2) was injected between the abdominal sternites of virgin *O. fasciatus* females. Maternal dsRNA injection has been used in *O. fasciatus* to post-transcriptionally knockdown expression in embryos developing from eggs laid by injected females[27] and is known to reduce expression of embryonically expressed genes. A control injection of double-stranded *Red* (ds-*red*) or buffer was used to confirm that there was no technical affect associated with RNAi treatment. The effects of RNAi treatment on gene expression was assessed in gut, head, thorax and ovary tissues 10 days after mating to untreated *O. fasciatus* males. The

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

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

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

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

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

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

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

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

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

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

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

References

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

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

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

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

Authors contributions. AJB, RJS, PJM, and AJM designed the study. ZS and PJM performed miscroscopy analyses. ECM performed RNAi, and extracted DNA and RNA for library prepation and sequencing. AJB analyzed the MethylC-seq and RNA-seq data, and wrote the manuscript with contributions from RJS, PJM, AJM and ECM.

Author information. The authors declare no competing interests.

Materials and correspondence. Correspondence and requests for materials should be addressed to pjmoore@uga.edu and schmitz@uga.edu.

Figure legends

Fig. 1: *Dnmt1* is required for reproduction in *O. fasciatus*. a, Assessment of RNAi treatment targeting *Dnmt1* using qRT-PCR demonstrates successful reduction in transcription in ovaries compared to control. Dots indicate mean expression level, and error bars indicate standard error of the mean. **b**, Parental RNAi injection with ds-*dnmt1* significantly affected the development of eggs laid by injected females compared to eggs laid by control females. Dots indicate mean expression level, and error bars indicate standard error of the mean. **c**, Whole ovaries from females removed 12-14 days post-injection (i and ii). In control females (i), mature oocytes can be seen collecting in the lateral oviduct (Od). In ds-*dnmt1* females (ii), no mature oocytes are apparent, but the oviduct has filled with a yolk-like substance. Scale bar equals 1 mm in (i) and (ii). High magnification of the follicular epithelium surrounding a maturing oocyte (iii and iv). Nuclei are stained with Hoechst 33258 and artificially colored turquoise. Nuclei from control follicular epithelium (iii) are round and regular in shape. Nuclei from ds-*dnmt1* females (iv) are highly irregular in shape. Scale bar equals 20 μm in (iii) and (iv).

Fig. 2: *Dnmt1* is required for mCG in *O. fasciatus*. a, Level and genomic location of mCG between hemi– and holometabolous insects, and *D. pulex*, and RNAi treatment targeting *Dnmt1* in *O. fasciatus*. **b**, Levels of mCG across gene bodies and one kilobase pairs (1kb) flanking sequence of hemi– and holometabolous insects, and *D. pulex*. Also shown are the levels of mCG across gene bodies for *O. fasciatus* ds-*dnmt1*. **c**, Density plot representation of mCG for RNAi treated *O. fasciatus* and species investigated in this study.

Fig. 3: Loss of mCG in *O. fasciatus* ovaries has a limited effect on transcription. a, Gene expression level for deciles of increasing mCG (1–10) and unmethylated genes (UM). Error bars represent 95% confidence interval of the mean. **b**, Regression of gene expression against a continuous measure of mCG with >0 FPKM for the same set of genes that are CG-methylated in *O. fasciatus* control, but unmethylated in *ds-dnmt1*. Raw *P* values are provided for each regression, and significance or non-significance (NS) is indicated in brackets following Bonferroni correction. **c**, Combinational overlap of genes that are differential CG-methylated and expressed, and similarly CG-methylated and expressed between *O. fasciatus ds-dnmt1* and control. Gene groups: Differentially Methylated Gene (DMG), Differentially Expressed Gene (DEG), Similarly Methylated Gene (SMG)/UnMethylated Genes (UMG), and non-Differentially Expressed Gene (non-DEG). **d**, A heatmap showing gene expression changes for genes that are differentially CG-methylated between *O. fasciatus* ds-*dnmt1* and control. Expression was standardized by the highest value per gene per biological replicate to produce a Relative Fragments Per Kilobase of transcript per Million (RFPKM) value. RFPKM were clustered using a hierarchical clustering method.

Fig. 4: Large-scale reactivation of TEs did not follow severe reductions of mCG in *O. fasciatus* ovaries. a, CG methylation levels for the top ten most abundant TEs of ≥500 bp in the *O. fasciatus* genome. The dashed lines correspond to the intergenic mCG level of *O. fasciatus* ds-*dnmt1* and control. **b**, Levels of mCG across the bodies and 1kb

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

Methods

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

PCR confirmation for the presence of a single *Dnmt1* ortholog in *O. fasciatus*. To determine if OFAS015351 and OFAS018396 were two parts of a single *Dnmt1* ortholog, we designed one sense primer at the 3' end of OFAS015351 (Of_DMNT1-1_3603S; Extended Data Table 1) and two antisense primers at the 3' end of OFAS018396 (Of_DNMT1-2_424A and Of_DNMT1-2_465A; Extended Data Table 1). A fourth sense primer (Of_DNMT1-2_1S; Extended Data Table 1) was designed at the 5' end of OFAS018396 to confirm the size and sequence of this possibly truncated gene annotation.

Polymerase Chain Reaction (PCR) with primer combinations Of_DMNT1-1_3603S–Of_DNMT1-2_424A, Of_DMNT1-1_3603S–Of_DNMT1-2_465A, and Of_DNMT1-2_1S–Of_DNMT1-2_424A was performed using Q5 Polymerase (New England BioLabs, Ipswich, MA) per manufactures instructions. Thermacycler conditions were 98°C for 15 seconds (s) (denaturing), 60°C for 30 s (annealing), and 72°C for 30 s (extension), and repeated for 40 cycles. The PCR products were then purified using QIAquick PCR Purification Kit (Qiagen, Venlo, The Netherlands), and sequenced at the Georgia Genomics and Bioinformatics Core (Athens, GA).

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

Parental RNAi. Template for the *in vitro* transcription of reactions was prepared from a PCR reaction in which T7 phage promoter sequences were added to the gene-specific *Dnmt1* primers[36]. For our control sequence, we used the red fluorescent protein (*Red*) sequence used in previous parental RNAi experiments in *O. fasciatus*[36] or buffer. Primer sequences can be found in Extended Data Table 2. Sense and anti-sense RNA was synthesized in a single reaction using the Ambion MEGAscript kit (ThermoFisher Sci, Waltham, MA). After purification, the double-stranded RNA (dsRNA) concentration was adjusted to 2 µg/µL in injection buffer (5 mM KCl, 0.1 mM NaH_2PO_4)[36]. Females were injected with 5 µL of dsRNA between the abdominal sternites using an insulin syringe. Following injection, females were paired with an un-injected male to stimulate oviposition and fertilize eggs. Individual females with their mate were housed in petri dishes with sunflower seeds, water and cotton wool as an oviposition site. The parental RNAi protocol has been reported to result in 100% penetrance by the third clutch of eggs[36] and this was also our experience.

Reproductive phenotype screening and analysis. Eggs were collected between days 4–10 post-injection and assessed for development. *Oncopeltus fasciatus* embryos change from a creamy white color to orange as they develop, which indicates viability. Thus, color change is a useful tool for assessing healthy development. We examined the number of developing eggs at 5 days post-oviposition, at which point viable eggs are clearly distinguishable from inviable eggs, as well as hatching rate of eggs from females injected with double-stranded *Dnmt1* (ds-*dnmt1*) and double-stranded *Red* (ds-*red*). The data for the number of eggs laid was normally distributed so differences among RNAi treated groups were tested using analysis of variance (ANOVA). The development data, however, was not normally distributed and consisted of binary states (developed and not developed), and so differences among treatment groups were analyzed with a Generalized Linear Model (GLM) using a Poisson distribution.

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

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

Expression level of *Dnmt1* was quantified by quantitative real-time PCR (qRT-PCR). Primers were designed for *Dnmt1* using the *O. fasciatus* genome as a reference (37). Actin and GAPDH were used as endogenous reference genes. Primer sequences can be found in Extended Data Table 2. We used Roche LightCycler 480 SYBR Green Master Mix with a Roche LightCycler 480 (Roche Applied Science, Indianapolis, IN) for qRT-PCR. All samples were run with 3 technical replicates using 10 μL reactions using the manufacturer's recommended protocol. Primer efficiency calculations, genomic contamination testing and endogenous control gene selection were performed as described by[37]. We used the ΔΔCT method[38] to examine differences in expression between ds-*dnmt1* and ds-*red* injected females[37].

Whole-Genome Bisulfite Sequencing (WGBS) and analysis of cytosine (DNA) methylation. MethylC-seq libraries for an *O. fasciatus* ds-*dnmt1* and control individual were prepared according to the protocol described in[39] using genomic DNA extracted from ovaries (see **Materials and Methods** section **Quantitative RT-PCR**). Libraries were single-end 75 bp sequenced on an Illumina NextSeq500 machine[40, 41]. Unmethylated lambda phage DNA was used to as a control for sodium bisulfite conversion, and an error rate of ~0.05% was estimated. *Oncopeltus fasciatus* ds-*dnmt1* and control were sequenced to a depth of ~18× and ~21×, which corresponded to an actual mapped coverage of ~9× and ~11×, respectively. Additionally, low pass (<1×) WGBS from gut, head, ovary, and thorax was performed for three control and ds-*dnmt1* biological replicates. *Blattella germanica* was additionally sequenced to generate equal numbers of hemi– and holometabolous insects investigated in this study. However, DNA was extracted from whole-body minus gastrointestinal tract. MethylC-seq libraries were prepared and sequenced identically to *O. fasciatus*. An error rate of ~0.14% was estimated from unmethylated lambda phage DNA. *Blattella germanica* was sequenced to a depth of ~8×, which corresponded to an actual mapped coverage of ~5×. WGBS data for *O. fasciatus* and *B. germanica* can be found on Gene Expression Omnibus (GEO) under accession GSE109199. Previously published WGBS data for *A. mellifera*[42], *Bo. mori*[17], *Daphnia pulex*[43], *N. vespilloides*[19], and *Z. nevadensis*[15] were downloaded from the Short Read Archive (SRA) using accessions SRR445803–4, SRR027157–9, SRR1552830, SRR2017555, and SRR3139749, respectively. Thus, DNA methylation was investigated for six insects from six different orders spread evenly across developmental groups, and a crustacean outgroup. WGBS data was aligned to each species respective genome assembly using the methylpy pipeline[44]. In brief, reads were trimmed of sequencing adapters using Cutadapt v1.9[45], and then mapped to both a converted forward strand (cytosines to thymines) and converted reverse strand (guanines

to adenines) using bowtie v1.1.1[46]. Reads that mapped to multiple locations, and clonal reads were removed.

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

Ortholog identification. Best BLASTp hit (arguments: -max_hsps 1 –max_target_seqs 1 evalue 1e-03) was used to identify orthologs between *O. fasciatus* and other insect species investigated.

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

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

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

503 settings were determined following best practices for DEG analysis as described by[54].
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[26] through combining annotations from Blast2Go PRO v4.1.9[55],
509 InterProScan v5.23-62.0 (arguments: -goterms -iprlookup -appl CDD,Pfam)[32], and
510 through sequence homology to *D. melanogaster* using BLASTp (arguments: -evalue
511 1.0e-03 -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[56] 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[26] using bowtie2 v2.2.9[57] 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[58]. 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**
538
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., Pfrender, 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 ifferential expression analysis of digital gene expression data. *Bioinformatics* **26**,
590 139–140 (2010).

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