EPIGENETICS IN SOCIAL INSECTS

A Dissertation
Presented to
The Academic Faculty

by

Karl M. Glastad

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in Biology in the
School of Biology

Georgia Institute of Technology
May 2016

EPIGENETICS IN SOCIAL INSECTS

Approved by:

Dr. Michael A. D. Goodisman, Advisor
School of Biology
Georgia Institute of Technology

Dr. I. King Jordan
School of Biology
Georgia Institute of Technology

Dr. Soojin V. Yi,
School of Biology
Georgia Institute of Technology

Dr. Nicole M. Gerardo
Department of Biology
Emory University

Dr. J. Todd Streelman
School of Biology
Georgia Institute of Technology

Date Approved: 12/10/2015

For Science.

ACKNOWLEDGEMENTS

I would like to thank my wife and closest friend Christina for making sacrifices both financially and personally in support of pursuit of a PhD. My advisor, Mike Goodisman was also integral to my successful completion of a PhD, having provided excellent advice and support, both emotionally and intellectually. Soojin Yi, a member of my committee, has also been very helpful towards furthering my academic career, having provided invaluable advice and serving as coauthor on multiple papers. I would also like to thank the other members of my committee for providing ideas and encouragement at committee meetings. Students in the Goodisman lab have also provided moral support and scientific insight throughout this process. In particular, I would like to thank Brendan Hunt for being highly-important during my early PhD towards developing my day-to-day scientific skills and ideas, as well as being a huge help towards developing hypotheses and formulating ideas throughout my graduate work. Without Brendan, it is unlikely I would be anywhere near the scientist I am today. Finally, I would like to thank my parents for being supportive throughout this process.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

CDS	Coding Sequence
ChIP	Chromatin Immunoprecipitation
CI	Confidence Interval
CpG	Cytosine Immediately followed by Guanine in 5’ to 3’ Direction
CpG o/e	Normalized CpG Dinucleotide Content
DBR	Differentially Bound Region
DMC	Differentially Methylated Cytosine
DMG	Differentially Methylated Gene
DMR	Differentially Methylated Region
dN	Nonsynonymous Substitution Rate
DNA	Deoxyribonucleic Acid
DNMT	DNA methyltransferase
dS	Synonymous Substitution Rate
FDR	False Discovery Rate
FPKM	Fragments per Kilobase of Exon per Million Mapped Fragments
GC Content	Guanine and Cytosine Content
GC content	Guanine and Cytosine Content
gDNA	Genomic DNA
GEO	Gene Expression Omnibus
GLM	Generalized Linear Model
GO	Gene Ontology
GpC	Guanine Immediately Followed by Cytosine in 5’ to 3’ Direction
H2A.Z	H2A Histone Variant
H3	Histone H3
H3K27ac	Acetylation of Lysine 27 of Histone H3

H3K27me3	Trimethylation of Lysine 27 of Histone H3
H3K36me3	Trimethylation of Lysine 36 of Histone H3
H3K4me1	Monomethylation of Lysine 4 of Histone 3
H3K4me3	Trimethylation of Lysine 4 of Histone 3
H3K9ac	Acetylation of Lysine 9 of Histone H3
H3K9me3	Trimethylation of Lysine 9 of Histone H3
HMR	Highly Methylated Region
hPTM	Histone Post Translational Modification
mCG	Methylated CpG Dinucleotide
miRNA	micro RNA
MLM	Multiple Linear Regression Model Coefficient
NCBI	National Center for Biotechnology Information
PAML	Phylogenetic Analysis by Maximum Likelihood
PC	Principal Component
PCA	Principal Component Analysis
R2	Linear Regression Coefficient of Determination
RNA	Ribonucleic Acid
RNA pol II	RNA polymerase II
TF	Transcription Factor
TFBS	Transcription Factor Binding Site
TSS	Transcription Start Site
TTS	Transcription Termination Site

SUMMARY

Virtually all multicellular organisms are capable of developing differently in response to environmental variation. Such environmental responsiveness is known as phenotypic plasticity, whereby a single genome can produce multiple distinct phenotypes based upon environmental information. Social insect castes are excellent examples of phenotypic plasticity, as the production of specialized castes is environmentally determined in most cases.

At the molecular level, phenotypic plasticity requires interpretation and perpetuation of environmental signals without changing the underlying genotype. Such non-genetic, heritable information is known as epigenetic information. This dissertation examines epigenetic information among social insects, and how differences in such information relate to phenotypic caste differences. The studies included herein primarily focus on one form of epigenetic information, DNA methylation. In particular, these studies explore DNA methylation as it relates to and impacts (i) alternative phenotype and gene expression differences, (ii) histone modifications, another important form of epigenetic information, in insect genomes, and (iii) molecular evolutionary rate of underlying actively transcribed gene sequences.

We find that DNA methylation exhibits marked epigenetic and evolutionary associations, and is linked with alternative phenotype in multiple insect species. Thus, DNA methylation is emerging as one important epigenetic mediator of phenotypic plasticity in social insects.

CHAPTER 1

INTRODUCTION

Environmental responsiveness plays a fundamental role in the success of complex life forms (West-Eberhard 2003; Pfennig, Wund et al. 2010). This is particularly evident in social insects, where the production of specialized castes has facilitated their ecological dominance (Wilson 1990). Importantly, castes often show extreme differences in morphology, physiology, and behavior arising through the differential expression of genes (Robinson, Grozinger et al. 2005; Smith, Toth et al. 2008). In the majority of social insect species castes are environmentally determined (Wheeler 1986), making them an excellent example of phenotypic plasticity, whereby a single genome can produce multiple distinct phenotypes based upon environmental differences/information (Evans and Wheeler 2001; West-Eberhard 2003).

Epigenetic information plays an important role in regulating the development of environmentally induced phenotypic variation (Kucharski, Maleszka et al. 2008; Burdge and Lillycrop 2010; Schmitz and Ecker 2012). Epigenetic information is simply any heritable information that can affect gene function that is not coded in the standard compliment of DNA bases (Berger, Kouzarides et al. 2009). Variation in epigenetic information can lead to sustained changes in gene expression (Kota and Feil 2010; Margueron and Reinberg 2010), ultimately permitting variation in developmental programs in response to environmental cues.

DNA methylation is one important epigenetic modification, and is present in all three domains of life (Klose and Bird 2006; Suzuki and Bird 2008; Zemach, McDaniel et al. 2010). DNA methylation has been implicated in the regulation of gene expression variation in mammals (Fraga, Ballestar et al. 2005; Cheong, Yamada et al. 2006; Reik 2007), plants (Li, Wang et al. 2008; He, Chen et al. 2011), and insects (Kucharski, Maleszka et al. 2008; Lyko, Foret et al. 2010; Glastad, Hunt et al. 2011). DNA methylation has also been linked to the outcome of alternative splicing (Lyko, Foret et al. 2010; Maunakea, Nagarajan et al. 2010; Shukla, Kavak et al. 2011), as well as chromatin structure and modification in both plants (Zhang, Bernatavichute et al. 2009; Chodavarapu, Feng et al. 2010) and vertebrates (Okitsu and Hsieh 2007; Hodges, Smith et al. 2009; Jeong, Liang et al. 2009). Thus DNA methylation appears to play an important role in mediating the relationship between genotype and phenotype in many taxa.

Notably, DNA methylation has been linked to caste formation in honeybees (Kucharski, Maleszka et al. 2008), and has also been identified in the genomes of several other hymenopteran social insect taxa (Glastad, Hunt et al. 2011). Caste-specific differences in DNA methylation are associated with alternative splicing differences between reproductive queens and sterile workers (Lyko, Foret et al. 2010), as well as between queen and worker-destined larvae in honey bees (Foret, Kucharski et al. 2012). Importantly, the mechanism through which DNA methylation differences impact the determination of caste, or how widespread this phenomenon is across insects, remains unknown.

Furthermore, DNA methylation has recently been identified in the genome of the termite *Coptotermes lacteus* (Lo, Li et al. 2012), and *Zootermopsis nevadensis* (Terrapon, Li et al. 2014). Termites are highly social and exhibit distinct castes (Scharf, Buckspan et al. 2007; Toru and Scharf 2011), but represent a completely novel origin of sociality from the Hymenoptera; isopteran and hymenopteran insects diverged approximately 375 MYA (Gaunt and Miles 2002). Thus, termites provide an important evolutionary contrast to Hymenoptera for investigating the link between developmental regulation of phenotypic plasticity and DNA methylation.

The research presented in this dissertation improves our understanding of DNA methylation across insects, providing evidence of caste-based DNA methylation differences in two economically-impactful social insect species. The analyses here further elucidate the evolutionary and epigenomic context of insect DNA methylation, and provide important insight into how this important epigenetic mark relates to social insect caste.

Chapter two of this dissertation (Glastad, Hunt et al. 2014) addresses the status of DNA methylation in the ant *Solenopsis invicta*, and how differences between phenotypes correspond to differences in DNA methylation. DNA methylation has recently been found to be an important regulator of caste in the honey bee (Kucharski, Maleszka et al. 2008; Lyko, Foret et al. 2010; Herb, Wolschin et al. 2012), and has been connected to regulating alternative splicing. This indicates DNA methylation may be important in the production of castes in ants. In our study of fire ant DNA methylation, we found that DNA methylation may be implicated in determining ant caste, and our study further suggests a role for DNA methylation in compensating for differences in ploidy between haplodiploid insect sexes.

Chapter three of this dissertation (Glastad, Liebig et al. in preparation) addresses DNA methylation in the termite *Z. nevadensis*. Termites represent an entirely distinct instance of sociality from Hymenoptera and are highly economically-impactful pests (Miura and Scharf 2011), yet there is a paucity of molecular data in this taxon. In this study, we present the first termite methylomes, and examine variation in methylation patterns among termite sexes and castes, in conjunction with paired gene expression data. We find that the termite genome possesses more DNA methylation than other studied insects, and that many genes are differentially methylated between termite castes. We also observe that differential methylation is associated with several functional regulatory motifs, potentially identifying an important mechanism through which DNA methylation impacts gene regulation.

Chapter four of this dissertation (Glastad, Hunt et al. 2015) addresses, for the first time in insects, the epigenomic context of DNA methylation. In mammals and plants, much work has been done to integrate DNA methylation into the broader epigenome, improving our understanding of both in the process. Importantly, studies in these model systems have shown that DNA methylation interacts with multiple chromatin modifications (Lorincz, Dickerson et al. 2004; Cedar and Bergman 2009). Until recently however, chromatin data has not existed for an insect with DNA methylation. In this study, we leverage recently independently-published DNA methylation (Bonasio, Li et al. 2012) and ChIP-seq (histone modification; (Simola, Ye et al. 2013)) data to elucidate the epigenetic and transcriptional context of DNA methylation in the Florida carpenter ant. We demonstrate that DNA methylation is targeted to specific chromatin regions of active genes, particularly those associated with the progression of RNA pol II from initiating to elongating forms. We further show that caste-specific differences in DNA methylation are significantly associated with caste-specific differences in several other key chromatin modifications between ant castes.

Chapter five of this dissertation addresses the molecular evolutionary implications of DNA methylation across several insects. Recent work in model systems has uncovered that the epigenome impacts molecular evolutionary rate (Tolstorukov, Volfovsky et al. 2011; Park, Qian et al. 2012), drawing an important link between evolution and genomic context. Here, we observe a strong association between DNA methylation and the neutral evolutionary rate of genes in hymenopteran insects. Surprisingly, we observe that this association is not entirely due to DNA methylation-associated mutagenesis, but instead seems to be associated with the active chromatin context that characterizes DNA methylation.

Overall, these studies greatly improve the phylogenetic breadth and genomic depth of our understanding of DNA methylation in social insects (and insects in general). Furthermore, we see that DNA methylation shows a complex association with alternative phenotype, and plays an important role in transcriptionally active, conserved insect gene bodies. This is evident from results from multiple phylogenetically-disparate social insect taxa, where (i) DNA methylation differs between alternative (but genomically highly-similar) phenotypes in termites and ants, (ii) is strongly associated with other important, transcription-associated epigenetic forms of information implicated in determining phenotype, and (iii) shows strong, consistent associations with evolutionary rate. Thus, DNA methylation appears to be a form of epigenetic information important to defining social insect alternative phenotype and contributing to more fundamental aspects of insect gene transcription.

CHAPTER 2

EPIGENETIC INHERITANCE AND GENOME REGULATION: IS DNA METHYLATION LINKED TO PLOIDY IN HAPLODIPLOID INSECTS?[1]

Abstract

Organisms show great variation in ploidy level. For example, chromosome copy number varies among cells, individuals and species. One particularly widespread example of ploidy variation is found in haplodiploid taxa, wherein males are typically haploid and females are typically diploid. Despite the prevalence of haplodiploidy, the regulatory consequences of having separate haploid and diploid genomes are poorly understood. In particular, it remains unknown whether epigenetic mechanisms contribute to regulatory compensation for genome dosage. To gain greater insight into the importance of epigenetic information to ploidy compensation, we examined DNA methylation differences among diploid queen, diploid worker, haploid male, and diploid male *Solenopsis invicta* fire ants. Surprisingly, we found that morphologically-dissimilar diploid males, queens, and workers were more similar to one another in terms of DNA methylation than were morphologically-similar haploid and diploid males. Moreover, methylation level was positively associated with gene expression for genes that were differentially methylated in haploid and diploid castes. These data demonstrate that intragenic DNA methylation levels differ among individuals of distinct ploidy and are positively associated with levels of gene expression. Thus, these results suggest that epigenetic information may be linked to ploidy compensation in haplodiploid insects. Overall, this study suggests that epigenetic mechanisms may be important to maintaining

[1] Glastad, K. M., B. G. Hunt, et al. 2014. Epigenetic inheritance and genome regulation: is DNA methylation linked to ploidy in haplodiploid insects? Proceedings of the Royal Society B: Biological Sciences 281.

appropriate patterns of gene regulation in biological systems that differ in genome copy number.

Introduction

Organisms display a remarkable diversity in ploidy level (Galitski, Saldanha et al. 1999; Edgar and Orr-Weaver 2001; Otto and Jarne 2001; Sassone-Corsi 2002; Heimpel and Boer 2008). For example, all sexual organisms show variation in ploidy during their life cycle. In addition, members of different species sometimes vary in ploidy number. Such ploidy variation shapes molecular evolution, genetic interactions, and gene function (Rasch, Cassidy et al. 1977; Galitski, Saldanha et al. 1999; Adams and Wendel 2005; Aron, de Menten et al. 2005; Aron, de Menten et al. 2005; Otto 2007). Thus, variation in ploidy fundamentally affects evolutionary and developmental processes.

A prime example of variation in ploidy is embodied by the haplodiploid genetic system. Haplodiploid species are typically characterized by having unfertilized eggs develop into haploid males and fertilized eggs develop into diploid females (Otto and Jarne 2001; Heimpel and Boer 2008). The haplodiploid genetic system has arisen at least 17 independent times during the course of animal evolution (Otto and Jarne 2001; Heimpel and Boer 2008), and is the ancestral genetic system of the order Hymenoptera (ants, bees, and wasps; Heimpel and Boer 2008; Heimpel and de Boer 2008). Consequently, as many as 20% of all animal species may be haplodiploid (Evans, Shearman et al. 2004; Evans, Shearman et al. 2004). Despite the taxonomic prevalence of haplodiploidy, the regulatory consequences of ploidy differences between sexes remain largely unknown (but see: Rasch, Cassidy et al. 1977; Aron, de Menten et al. 2005; Scholes, Suarez et al. 2013). This lack of information represents a gap in our understanding of how biological systems respond to ploidy variation.

Epigenetic modifications to chromatin are prime candidates for regulating gene function in haplodiploid taxa. Epigenetic marks are heritable and make fundamental contributions to gene regulation (Bonasio, Tu et al. 2010). One of the most important types of epigenetic marks is the methylation of DNA. DNA methylation is found in all three domains of life, suggesting a role in the common ancestor of all Metazoa (Klose and Bird 2006; Suzuki and Bird 2008).

Recently, DNA methylation and histone modifications have been implicated in the regulation of social insect caste differences (Kucharski, Maleszka et al. 2008; Lyko, Foret et al. 2010; Bonasio, Li et al. 2012; Foret, Kucharski et al. 2012; Simola, Ye et al. 2013). In addition, global sex chromosome dosage compensation is achieved in *Drosophila* and mammals by epigenetic mechanisms (Payer and Lee 2008; Conrad and Akhtar 2012), demonstrating that distinct epigenetic states can achieve transcriptional compensation associated with ploidy variation. However, the contributions of epigenetic inheritance to regulatory mechanisms that compensate for ploidy differences in haplodiploids have not been investigated. In this study, we attempted to gain insight into whether epigenetic information was associated with gene regulation in haplodiploid taxa.

In order to assess the epigenetic states of haploid and diploid genomes, we compared single nucleotide resolution DNA methylation profiles (DNA methylomes) of haploid and diploid individuals of the red imported fire ant, *Solenopsis invicta*. Sex in *S. invicta*, and many other hymenopteran insects, is determined by complementary sex determination (Heimpel and Boer 2008). Under single-locus complementary sex determination, sex is controlled by zygosity at a single genetic locus. In this case, heterozygous individuals develop into females and hemizygous (haploid) individuals develop into males.

Interestingly, diploid individuals that are homozygous at the sex determining locus develop into diploid males. Diploid males are generally rare in hymenopteran populations. However, diploid males are produced at high frequency in invasive *S. invicta* due to loss of variation at the sex-determining locus (Ross, Vargo et al. 1993; Ross, Vargo et al. 1993). *S. invicta* diploid males are larger than haploid males but otherwise have highly similar morphologies and behaviors to haploid males. Moreover, haploid and diploid males differ substantially in phenotype from diploid queens and workers (Ross and Fletcher 1985; Krieger, Ross et al. 1999). Importantly, the common production of haploid and diploid males makes *S. invicta* well-suited to investigate epigenetic gene regulation in the context of ploidy differences while simultaneously controlling for sex differences.

Our analyses uncovered striking differences in DNA methylation between haploid and diploid individuals in *S. invicta*. The link between DNA methylation and ploidy variation suggests that haploid and diploid genomes in *S. invicta* exhibit distinct epigenetic states. These results provide support for the hypothesis that epigenetic mechanisms are associated with genomic dosage compensation of haplodiploid organisms. More broadly, our results suggest that epigenetic information may influence the evolution of ploidy differences among cells, organisms, and species.

Material and Methods

Whole-genome bisulfite-sequencing

Sample collection, DNA extraction, bisulfite conversion, sequencing, quality control, and read mapping were performed as described elsewhere (Hunt, Glastad et al. 2013). Briefly, all samples were taken from a single *S. invicta* colony. Male ploidy was confirmed by DNA microsatellite analysis at 3-4 highly variable loci. Genomic DNA was separately pooled from whole bodies of haploid males, diploid males, alate queens, and workers, comprising one sample per caste. We obtained between 7-9x mean coverage of genomic CpG sites per sample (Hunt, Glastad et al. 2013). *S. invicta* SI2.2.3 gene models were used for analysis of genes, exons, and introns (Wurm, Wang et al. 2011). *S. invicta* whole-genome bisulfite-sequencing data are available online from Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/; GSE39959).

DNA methylation targets and levels

Significantly methylated CpG sites were assessed using a binomial test, implemented using the Math::CDF module in Perl, which incorporated deamination rate (from our unmethylated control) as the probability of success, and assigned a significance value to each CpG site related to the number of unconverted reads (putatively methylated Cs) as they compare to the expected number from control (Lyko, Foret et al. 2010). Resulting *P*-values were then adjusted for multiple testing (Benjamini and Hochberg 1995). Only sites with false discovery rate (FDR) corrected binomial *P* values < 0.01 and ≥ 3 reads were considered "methylated". Fractional methylation values were calculated, as described previously (Hunt, Glastad et al. 2013; Hunt, Glastad et al. 2013), for each CpG site or for each genomic feature (exons and introns).

Hierarchical clustering and dendrogram generation

The pvclust package in the R statistical computing environment was used to generate clustering and dendrogram diagrams of fractional methylation values of exons and introns (R Development Core Team 2011; R Development Core Team 2013). We used the "average" linkage agglomeration method, the "correlation" distance measure, and 1000 bootstrap replications. Only those genomic features (exons and introns) targeted by DNA methylation in at least one caste, according to FDR-corrected binomial tests, were included in hierarchical clustering analysis. Fractional DNA methylation values of a given exon or intron in castes that did not exhibit significant DNA methylation were set to zero prior to hierarchical clustering in order to minimize noise contributed by unconverted, putatively unmethylated cytosines.

Differential DNA methylation

Significantly differentially methylated features (exons and introns) were assessed for each pairwise comparison between castes using generalized linear models (GLM), implemented in the R statistical computing environment (R Development Core Team 2011; R Development Core Team 2013), where methylation levels for features were modeled as functions of "caste" and "CpG position". If caste contributed significantly (chi-square test of GLM terms) to the methylation status of a feature (after adjustment for multiple testing using the method of (Benjamini and Hochberg 1995)), it was considered differentially methylated between castes (Lyko, Foret et al. 2010). Only CpG sites that were methylated in one or both castes and covered by ≥ 4 reads in both libraries were used in these comparisons, and only features with ≥ 3 such CpG sites were considered in further analyses.

Once exons and introns were assigned differential methylation status using the above GLM, each significantly differentially methylated exon or intron was called as elevated in the caste with higher fractional methylation status of that feature. These features were then combined by gene, and each gene was called as a unidirectional differentially methylated gene if greater than two-thirds of the gene's differentially methylated features were elevated in the same direction.

Gene ontology

Gene ontology (GO) annotations were assigned using Blast2GO (Conesa, Gotz et al. 2005). Significant enrichment was assessed with a Fisher's exact test and corrected for multiple testing with a Benjamini–Hochberg false discovery rate (FDR) (Benjamini and Hochberg 1995). The "generic GO slim" subset of GO terms was used to assess significantly enriched terms (FDR, $P < 0.05$).

Gene expression

S. invicta whole-body cDNA microarray data (Wang, Jemielity et al. 2007; Wang, Jemielity et al. 2007; Ometto, Shoemaker et al. 2011) were mapped to *S. invicta* gene models as described previously (Hunt, Glastad et al. 2013; Hunt, Glastad et al. 2013). Expression ratios between queen, worker, and haploid male castes (Ometto, Shoemaker et al. 2011) were calculated as $\log_2\left(\left(\overline{\mathrm{C1}_{pupa} + \mathrm{C1}_{adult}}\right)/\left(\overline{\mathrm{C2}_{pupa} + \mathrm{C2}_{adult}}\right)\right)$, where C1 is the expression value estimated by BAGEL (Townsend and Hartl 2002; Townsend and Hartl 2002) for the first caste and C2 is the estimated expression value for the second caste.

For each gene, we assessed the coefficient of variation (standard deviation / mean; CV) of expression values as the mean of CV values calculated separately for whole body *S. invicta* adult and pupal workers, queens, and haploid males (median of 5 biological replicates per morph) (Ometto, Shoemaker et al. 2011).

For array data from haploid and diploid males (Gene Expression Omnibus accession: GSE42786 and GSE35217 (Wang, Wurm et al. 2013; Nipitwattanaphon, Wang et al. Forthcoming.), the Limma R package (Smyth 2005) was used to perform background correction (method=”normexp”), within- and between-array normalization (method=”printtiploess” and method=”Rquantile” respectively), followed by generation of gene expression ratios between haploid and diploid male arrays.

Coding sequence evolution

We used OrthoDB (Waterhouse, Zdobnov et al. 2011) 12-insect orthology data to assign single-copy orthologs between the ants *S. invicta*, *Pogonomyrmex barbatus*, and *Linepithema humile*. Nonsynonymous substitutions per nonsynonymous site and synonymous substitutions per synonymous site were determined for the *S. invicta* lineage using codeml in PAML as described previously (Hunt, Ometto et al. 2011). Genes with aligned sequence length ≤ 100, dS ≥ 4, or dN/dS ≥ 4 were filtered out prior to analysis.

Results

DNA methylation is associated with ploidy in *S. invicta*

We observed significant differences in methylation level in one or more pairwise comparison between castes for 3,478 exons (32.7% of 10,628 exons methylated in one or more caste) and 577 introns (23.3% of 2,479 introns methylated in one or more caste) in *S. invicta*. Ultimately, we classified any gene with a significant difference in the methylation level of at least one exon or intron, in at least one pairwise comparison between castes, as a differentially methylated gene (DMG).

We found that DNA methylation levels in all libraries derived from diploid individuals were more similar to one another than to the library derived from haploid males (Figure 2.1a). Diploid males, queens, and workers all showed methylation profiles that were highly diverged from haploid males. In particular, the majority of significantly differential methylation occurred between the haploid and diploid castes (Figure 2.1b, Figure A.1). The pairwise comparison with the greatest number of DMGs was that between haploid and diploid males. This is particularly noteworthy given the high degree of morphological and behavioral similarity between haploid and diploid males in *S. invicta* (Ross and Fletcher 1985; Krieger, Ross et al. 1999). The pairwise comparison with the fewest differences was that between queens and workers, both of which are diploid females (Figure 2.1b, Figure A.1). We note that these findings are unlikely the result of bisulfite conversion efficiency, as the queen library exhibited the highest unmethylated cytosine non-conversion rate, and haploid and diploid males had the most similar unmethylated cytosine nonconversion rate among all libraries (Table A.1).

We next defined directional DMGs as those wherein at least two-thirds of differentially methylated features (exons and introns) were more highly methylated in one caste of a given pairwise comparison. For example, if three of four differentially methylated features were more highly methylated in haploid males, then the gene would be categorized as having elevated methylation in haploid males relative to diploid males. In contrast, if two of four differentially methylated features were more highly methylated in haploid males (with the other two more highly methylated in diploid males), then the gene would not be characterized as a directional DMG. Analysis of directional DMGs provided insight into the castes that most frequently exhibited elevated DNA methylation levels. In each comparison between haploid and diploid castes, we observed considerably

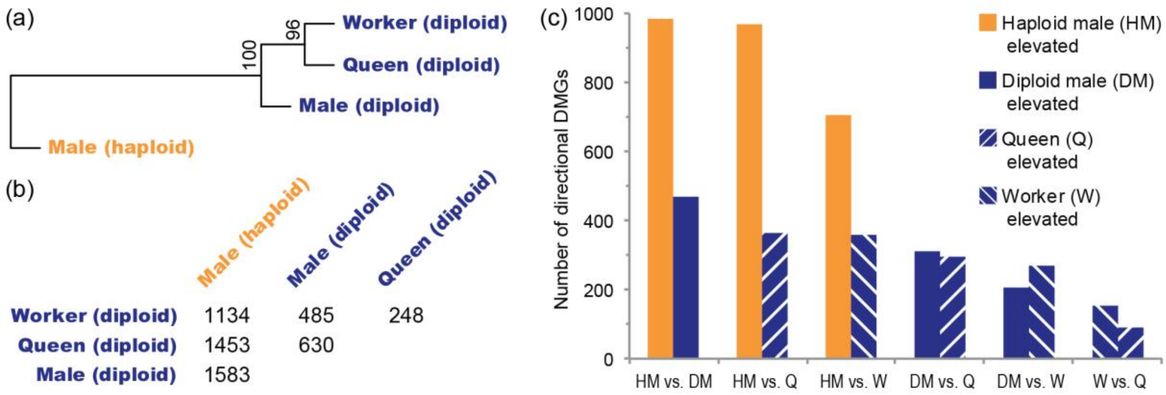


Figure 2.1. DNA methylation differs between haploid and diploid castes in *S. invicta*. (a) Dendrogram produced by hierarchical clustering of fractional methylation levels representing all introns and exons targeted by DNA methylation in at least one library (n = 10,560 genetic features); bootstrap probability values are shown. (b) Number of differentially methylated genes (DMGs) detected between castes. (c) Number of directional DMGs from panel b that exhibit pairwise elevated methylation in haploid (orange) and diploid (blue) castes, respectively.

more DMGs with elevated methylation levels biased to the haploid caste (Figure 2.1c, Figure A.1).

Differentially methylated genes in *S. invicta* have unique characteristics

We conducted enrichment analysis of gene ontology annotations for DMGs relative to methylated non-DMGs. We found that DMGs in *S. invicta* were enriched for annotations including “nucleotide binding” and “developmental process” (Table 1.1, Table A.2). In contrast, non-DMGs were enriched for terms related to core cellular functions such as “translation” (Table 2.1, Table A.3), as is typical of methylated genes in general in *S. invicta* and other insects (Glastad, Hunt et al. 2011; Hunt, Glastad et al. 2013; Hunt, Glastad et al. 2013).

We further tested whether there were significant differences between DMGs and non-DMGs in a number of gene characteristics in order to better understand which types of genes are variably methylated. Specifically, we determined if DMGs and non-DMGs differed in overall DNA methylation level (all castes combined), gene length, gene expression variability among samples as measured by the coefficient of variation (Ometto, Shoemaker et al. 2011), and rates of protein coding sequence evolution.

We found that DMGs exhibited substantially lower DNA methylation levels, and were substantially longer in terms of both coding sequence and gene body, than non-DMGs (Table 2.2, $P < 0.0001$ in each case). DMGs were also modestly, but significantly, more variable in expression, and more highly conserved at the sequence level, than non-DMGs (Table 2.2, $P < 0.01$ in each case).

We next investigated if variation in DNA methylation was associated with variation in gene expression among castes. In order to investigate the regulatory significance of differential DNA methylation, we integrated available microarray gene expression data from *S. invicta* haploid males, diploid queens, and diploid workers (Ometto, Shoemaker et al. 2011), as well as from a separate comparison of haploid and diploid males (Wang, Wurm et al. 2013; Nipitwattanaphon, Wang et al. Forthcoming.). Our analyses revealed that directional DMGs with elevated methylation in haploid castes versus diploid castes were significantly more highly expressed in haploid castes than in diploid castes (Figure 2.2, Figure A.1). This finding is consistent with the observed association between DNA methylation and active gene expression in insects (Foret, Kucharski et al. 2009; Hunt, Glastad et al. 2013; Hunt, Glastad et al. 2013). Intriguingly, however, we found no significant association between differential methylation and gene expression bias when examining genes differentially methylated between worker and queen castes (both diploid; Figure 2.2b).

Finally, we determined whether directional DMGs between males of different ploidy were enriched for distinct gene ontology annotations. Our goal with this analysis was to determine if elevated methylation in haploid males, which may reflect an epigenetic state associated with haploid gene upregulation (Figure 2.2a and Figure 2.2b), was targeted to genes associated with distinct functions, as compared to other DMGs.

Table 2.1. Enrichment of gene ontology (GO) annotations associated with differentially methylated genes (DMGs) and non-DMGs in *S. invicta*

Term	Category[a]	Accession	Number of genes	Fold enrichment[b]	FDR *P*-value
Differentially methylated genes					
Binding	F	GO:0005488	1434	1.13	0.0001
nucleotide binding	F	GO:0000166	464	1.28	0.0277
developmental process	P	GO:0032502	278	1.36	0.0435
Chromosome	C	GO:0005694	122	1.67	0.0435
Non-differentially methylated genes					
structural molecule activity	F	GO:0005488	64	1.92	0.0082
cytoplasmic part	C	GO:0000166	447	1.26	0.0103
Ribosome	C	GO:0032502	111	1.51	0.0427
Translation	P	GO:0005694	138	1.42	0.0488

[a] P, biological process; F, molecular function; C, cellular component

[b] Enrichment determined for DMGs or non-DMGs relative to all other methylated genes

We found that genes with elevated methylation in haploid males relative to diploid males were enriched for several metabolic process terms, as well as the terms “nucleotide binding” and “chromosome” (Table A.4). In contrast, there were no significantly enriched terms below the false discovery rate cutoff (FDR $P < 0.05$) for genes with elevated methylation in diploid males relative to haploid males. Nevertheless, several terms related to growth and development, including “developmental process”, were enriched among genes with elevated methylation in diploid males prior to FDR correction ($P < 0.05$; Table A.5). Together, these data suggest a marked difference between the gene classes that exhibit elevated methylation in haploid and diploid males. Elevated methylation in haploid males appears to preferentially target genes associated with basal cellular processes, whereas elevated methylation in diploid males may be associated with a larger number of genes implicated in development.

Orthologs of genes implicated in *Drosophila* dosage compensation exhibit differential DNA methylation and expression in *S. invicta*

We assessed patterns of DNA methylation and gene expression for *S. invicta* orthologs of genes associated with dosage compensation in *Drosophila.* Our goal was to provide initial insight into whether common molecular machinery may underlie dosage compensation for sex chromosomes in *Drosophila* and regulatory compensation for haploidy versus diploidy in *S. invicta*. Interestingly, we found that orthologs of four of eight genes (with data) related to dosage compensation in *D. melanogaster* were differentially methylated between haploid and diploid males in *S. invicta* (Table A.6, Figure A.2). Moreover, three of four of these genes (with data) were differentially expressed between haploids and diploids (Table A.6). Thus, several genes involved in *Drosophila* dosage compensation are differentially methylated and differentially expressed between haploids and diploids in *S. invicta*.

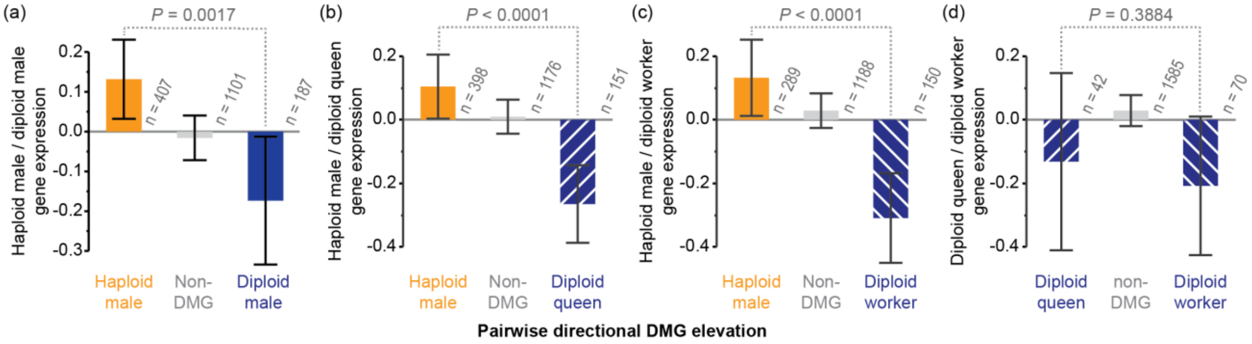


Figure 2.2. Gene expression bias is associated with directional differentially methylated genes (DMGs) in *S. invicta*. DMGs that exhibit elevated methylation in haploid males are more highly expressed in haploid males, whereas DMGs that exhibit elevated methylation in diploid (a) males, (b) queens, or (c) workers are more highly expressed in diploid males, queens, or workers, respectively. In contrast, there is no significant difference between the ratio of expression for DMGs that exhibit elevated methylation in (d) a pairwise comparison of queens and workers. Expression ratio data were standardized (mean zero, unit variance) following $\log_2$-transformation; *P*-values denote the results of Mann-Whitney U tests.

Discussion

The purpose of this investigation was to gain a greater understanding of the molecular mechanisms that regulate gene function among individuals that differ in ploidy. Our analysis of DNA methylation patterns among *S. invicta* castes uncovered strong associations between levels of DNA methylation and ploidy. These methylation differences were further found to be related to gene expression differences among castes.

We found most DMGs arose between haploid and diploid castes, and, therefore, that the number of DMGs was not related to the overall morphological similarity of the castes being compared (Figure 2.1). In particular, our comparison of haploid and diploid males produced more DMGs than our comparison of haploid males and diploid queens, which are sexually dimorphic, and produced many more DMGs than were observed between diploid queens and diploid workers, which are a classical example of insect polyphenism. Thus, in *S. invicta*, differences in DNA methylation more closely track differences in ploidy than differences in morphology, behavior, or physiology associated with distinct queen and worker castes (Smith, Toth et al. 2008).

The DNA methylomes of haploid males and diploid females were sequenced previously in the ants *Camponotus floridanus* and *Harpegnathos saltator* (Bonasio, Li et al. 2012; Bonasio *et al.* 2012). When we assessed directional DMGs between adult castes of *C. floridanus* and *H. saltator*, we found that, in four of six comparisons between haploid and diploid castes, more DMGs were elevated in haploids than in diploids (three of three comparisons in *H. saltator* and one of three comparisons in *C. floridanus*; Figure A.3). Thus, the data of Bonasio et al. further suggest that haploids may be prone to elevated DNA methylation relative to diploids.

Intriguingly, we found that differentially methylated genes (DMGs), as a whole, exhibited several distinguishing characteristics in *S. invicta*. DMGs were enriched relative to non-DMGs for the gene ontology annotations "nucleotide binding" and "developmental process" (Table 2.1), consistent with important regulatory roles for differential DNA methylation in *S. invicta*, as in the honey bee (Kucharski, Maleszka et al. 2008; Lyko, Foret et al. 2010; Foret, Kucharski et al. 2012; Foret, Kucharski et al. 2012). Furthermore, DMGs differed significantly from other methylated genes in methylation level, gene length, expression variability, and substitution rate (Table 2.2), suggesting key architectural and regulatory differences between DMGs and non-DMGs.

In *S. invicta*, differential methylation events were also associated with ploidy-specific gene expression bias (Figure 2.2), suggesting that DMGs are associated with regulatory differences between haploid and diploid genomes. Interestingly, the association of intragenic DNA methylation with active gene expression in insects suggests DNA methylation may be a useful marker of active chromatin states (Foret, Kucharski et al. 2009; Glastad, Hunt et al. 2011; Hunt, Glastad et al. 2013). In support of this idea, the presence of DNA methylation has recently been linked to the presence of

Table 2.2. Characteristics of differentially methylated genes (DMGs) and non-DMGs in *S. invicta*

Trait	**DMGs (mean ± SE; gene count)**	**Non-DMGs (mean ± SE; gene count)**	**Mann-Whitney U test *P*-value**
Fractional coding sequence methylation	0.186 ± 0.003; 2518	0.255 ± 0.005; 1705	< 0.0001
Fractional gene body (exons + introns) methylation	0.174 ± 0.002; 2521	0.240 ± 0.004; 1718	< 0.0001
Coding sequence length	1888.09 ± 31.07; 2518	1254.34 ± 29.27; 1705	< 0.0001
Gene body length	3869.21 ± 74.75; 2521	2483.83 ± 67.44; 1718	< 0.0001
Gene expression coefficient of variation	0.167 ± 0.002; 1068	0.160 ± 0.002; 836	0.0022
Nonsynonymous substitutions per nonsynonymous site (dN)	0.033 ± 0.001; 1728	0.037 ± 0.001; 1012	0.0046
Synonymous substitutions per synonymous site (dS)	0.344 ± 0.004; 1727	0.375 ± 0.005; 1012	< 0.0001
dN/dS	0.101 ± 0.002; 1725	0.102 ± 0.003; 1012	0.7962

several active histone modifications in insect genomes (Nanty, Carbajosa et al. 2011; Hunt, Glastad et al. 2013; Hunt, Glastad et al. 2013). We speculate that elevated haploid DNA methylation may be indicative of regulatory pressures associated with the single copy state of haploid loci.

Notably, our data cannot directly address whether changes in DNA methylation are the cause or consequence of changes in gene expression. However, experimental investigations in model systems indicate the DNA methylation can cause changes in gene function through interactions with other components of chromatin. For example, DNA methylation has been shown to affect alternative splicing through its interaction with RNA polymerase II (Shukla, Kavak et al. 2011). In addition, DNA methylation has been shown to alter the positioning of certain histone variants, which ultimately influence gene expression (Zilberman, Coleman-Derr et al. 2008). Experimental changes in levels of DNA methylation have also been found to lead to changes in levels of gene expression in *Arabidopsis* (Zilberman *et al.* 2008; Zilberman *et al.* 2007) , suggesting that intragenic methylation has functional effects.

The suggestion that epigenetic gene regulation plays a role in genome-wide chromosomal dosage compensation is consistent with the observation that epigenetic marks play key roles in sex chromosome dosage compensation (Payer and Lee 2008; Gelbart and Kuroda 2009; Conrad and Akhtar 2012). Intriguingly, we found that *S. invicta* orthologs of several genes implicated in *D. melanogaster* dosage compensation were differentially regulated between haploid and diploid castes (Table A.6, Figure A.2), raising the prospect for some degree of molecular convergence. Although the genome of *D. melanogaster* is not substantially methylated, previous studies have revealed that, in species that harbor functional DNA methylation systems, DNA methylation interacts with histone modifications associated with dosage compensation in *D. melanogaster* (Nanty, Carbajosa et al. 2011; Hunt, Glastad et al. 2013). Regardless, we note that the mechanisms by which intragenic methylation affect gene function remain poorly understood (Hunt, Glastad et al. 2013; Hunt, Glastad et al. 2013), and direct connections between mechanisms of sex chromosome dosage compensation and ploidy compensation remain speculative at present.

Given the evidence for different epigenetic states in haploid and diploid *S. invicta*, it is important to consider why one may expect different regulatory requirements for genes in haploid genomes as compared to diploid genomes. For example, there may be increased metabolic requirements placed on loci in haploid, relative to diploid, genomes (Edgar and Orr-Weaver 2001). Our results agree with this notion, as several metabolic process gene ontology annotations were enriched among genes with elevated DNA methylation in haploid males (Table A.4). One additional reason for epigenetic states to differ between haploid and diploid genomes may be related to the amelioration of haploid gene expression noise, particularly at genes essential to cellular function. Indeed, gene expression variability is negatively associated with dosage in yeast, where diploid cells exhibit less expression variability than haploid cells (Wang and Zhang 2011a; Wang and Zhang 2011b) , and where overall gene expression variability can lower organismal fitness (Talia, Skotheim et al. 2007). We previously found that DNA methylation is negatively associated with the coefficient of variation of gene expression among replicate *S. invicta* samples (Hunt, Glastad et al. 2013; Hunt, Glastad et al. 2013), potentially implying a role for DNA methylation in the stabilization of gene expression (Huh, Zeng et al. 2013). We speculate that, if DNA methylation plays a role in reducing gene expression stochasticity (Huh, Zeng et al. 2013), the variable expression of haploid loci may itself provide an impetus for elevated levels of DNA methylation in haploid males.

Overall, our results suggest that epigenetic mechanisms are associated with regulatory response to global differences in dosage in haplodiploid hymenopterans. However, we must emphasize that these results are preliminary in nature, requiring additional study to resolve whether epigenetic information is functionally implicated in ploidy-associated regulatory compensation. One important consideration is that haploid males in Hymenoptera are known to compensate for lower genomic content relative to diploid females through endoreplication (Rasch, Cassidy et al. 1977; Aron, de Menten et al. 2005), wherein cells increase their genomic content without dividing (Edgar and Orr-Weaver 2001). Our results raise the possibility that epigenetic information similarly contributes to haploid regulatory compensation, particularly given that endoreplication is not ubiquitous among tissues (Aron, de Menten et al. 2005). An alternative, but presently unexplored, possibility is that endoreplication itself is associated with epigenetic changes.

We have shown that differential DNA methylation is more closely linked to ploidy variation than to queen and worker castes in the fire ant *S. invicta*. We observed elevated DNA methylation in haploids and a positive association between ploidy-biased DNA methylation and gene expression, which together demonstrate the existence of distinct epigenetic states for haploid and diploid genomes. Overall, our results highlight the prospect that epigenetic mechanisms may be involved in achieving ploidy compensation in haplodiploid taxa.

Acknowledgments

This work was supported by the U. S. National Science Foundation (grant numbers DEB-1011349, DEB-0640690, IOS-0821130, and MCB-0950896) and the Georgia Tech-Elizabeth Smithgall Watts endowment.

CHAPTER 3

THE CASTE- AND SEX-SPECIFIC DNA METHYLOME OF THE TERMITE *ZOOTERMOPSIS NEVADENSIS*[2]

Abstract

DNA methylation is a widely-conserved epigenetic signal that has recently been demonstrated to play an important role in mediating alternative phenotype in social insects. To date, studies of DNA methylation have been confined to hymenopteran social insects, despite multiple non-hymenopteran origins of sociality among insects, such as seen in the termites, which have recently been identified as having considerable levels of DNA methylation. In order to extend our understanding of social insect epigenetics to a diverse yet poorly studied clade of independently eusocial insects, we performed replicated bisulfite and transcriptome sequencing of both sexes among multiple castes of the termite *Zootermopsis nevadensis*, for which a genome was recently published. We find some of the highest levels of DNA methylation found to date in an insect, as well as strong evidence of caste-associated differential methylation independent of sex differences (which were minimal). Differentially methylated genes were also more likely to be alternatively spliced than non-differentially methylated genes. We further observed strong functional enrichment of differentially methylated genes, suggesting a yet-unobserved regulatory function of DNA methylation in the production of termite castes. We further provide one potential such mechanism, observing strong overrepresentation of multiple TFBS and miRNA profiles within DMRs, many of which show preferential association with caste- or sex-specific differential methylation. Overall, our results show that DNA methylation is widespread and associated with caste in termites, and more generally provide important evolutionary insights into the relationship between DNA

[2] Glastad, K. M., J. Liebig, et al. in preparation. The caste- and sex-specific DNA methylome of the termite *Zootermopsis nevadensis*.

methylation and insect alternative phenotype. Furthermore our results suggest that termites represent an excellent, evolutionarily distinct alternative to hymenopteran social insects for studies of the molecular basis of caste.

Introduction

Phenotypic plasticity is a highly important mechanism, whereby a single genotype can produce multiple phenotypes based upon environment. Social insects represent excellent models for studies of phenotypic plasticity. In social insect societies, highly related individuals often develop distinct phenotypes, usually through the integration of information from the environment (West-Eberhard 2003). At its core, such phenotypic plasticity requires epigenetic information. Epigenetic information is any information not coded in the standard compliment of DNA bases, that nevertheless effects/encodes heritable changes in gene expression (Berger, Kouzarides et al. 2009).

Recently, DNA methylation, one form of epigenetic information, has been implicated as an important component of the determination of caste in at least one social insect. Indeed, DNA methylation was shown to have a direct impact on the production of castes in the honey bee (Kucharski, Maleszka et al. 2008). DNA methylation has further been associated with alternative splicing differences between honey bee castes (Lyko, Foret et al. 2010; Herb, Wolschin et al. 2012), and has been found to differ between castes in other hymenopteran social insects(Bonasio, Li et al. 2012; Glastad, Hunt et al. 2014). Importantly, while differences in DNA methylation have been observed between castes of multiple hymenopteran social insects, the exact mechanisms whereby DNA methylation effects developmental plasticity (if indeed it does) remain to be elucidated.

While much work has been done in Hymenopteran social insects to evaluate and explore the molecular bases of caste, this is not the only social insect. Termites represent an entirely novel origin of eusociality, and are distinct from the hymenoptera in many ways (Eggleton 2011). Being a close relative of wood-dwelling roaches, termites possess hemimetabolous-based caste system distinct from that seen in the holometabolous hymenopteran social insects. Thus, the developmental program that underlies termite societies differs substantially from the hymenoptera. In lower termites for example, workers, while considered a distinct caste, are composed of multiple, developmentally progressive instars. Furthermore, mature worker instars are poised to develop into either soldiers, winged reproductives (through an intermediate nymph stage), and in some species, a specialized worker-derived (wingless) reproductive form (Eggleton 2011).

Much like in hymenopteran social insects, this developmental plasticity is largely informed by hormonal (endogenous) and environmental (exogenous) cues (Mao, Henderson et al. 2005; Scharf, Buckspan et al. 2007; Toru and Scharf 2011). Unlike in hymenopteran social insects however, development and castes are arguably more protean. Termites further differ from hymenopteran social insects in that both sexes are near-equally represented among the majority of castes (Eggleton 2011), allowing for an examination of caste differences whilst controlling for sex. Thus, termites are an enticing system for studying the molecular basis of caste, and provide an under-studied alternative to the hymenopteran social insects.

Notably, preliminary research indicates that termites possess a functional suite of DNA methyltransferase enzymes, as well as putative DNA methylation in their genome (Terrapon, Li et al. 2014). Here, we present the first DNA methylomes from the termite, *Zootermopsis nevadensis*. We performed replicated BS-seq and stranded RNA-seq for both sexes of two termite castes (male and female alates and workers). We find that DNA methylation is considerably higher and targets more genes in the termite genome than seen in other social insects, resulting in large regions where most CpGs are methylated -- corresponding to gene-dense regions. Furthermore, we find many differences in DNA methylation between our morphs, the great majority of which differ between castes. While we find that genes containing differential methylation are functionally enriched for multiple development-associated processes, they are actually

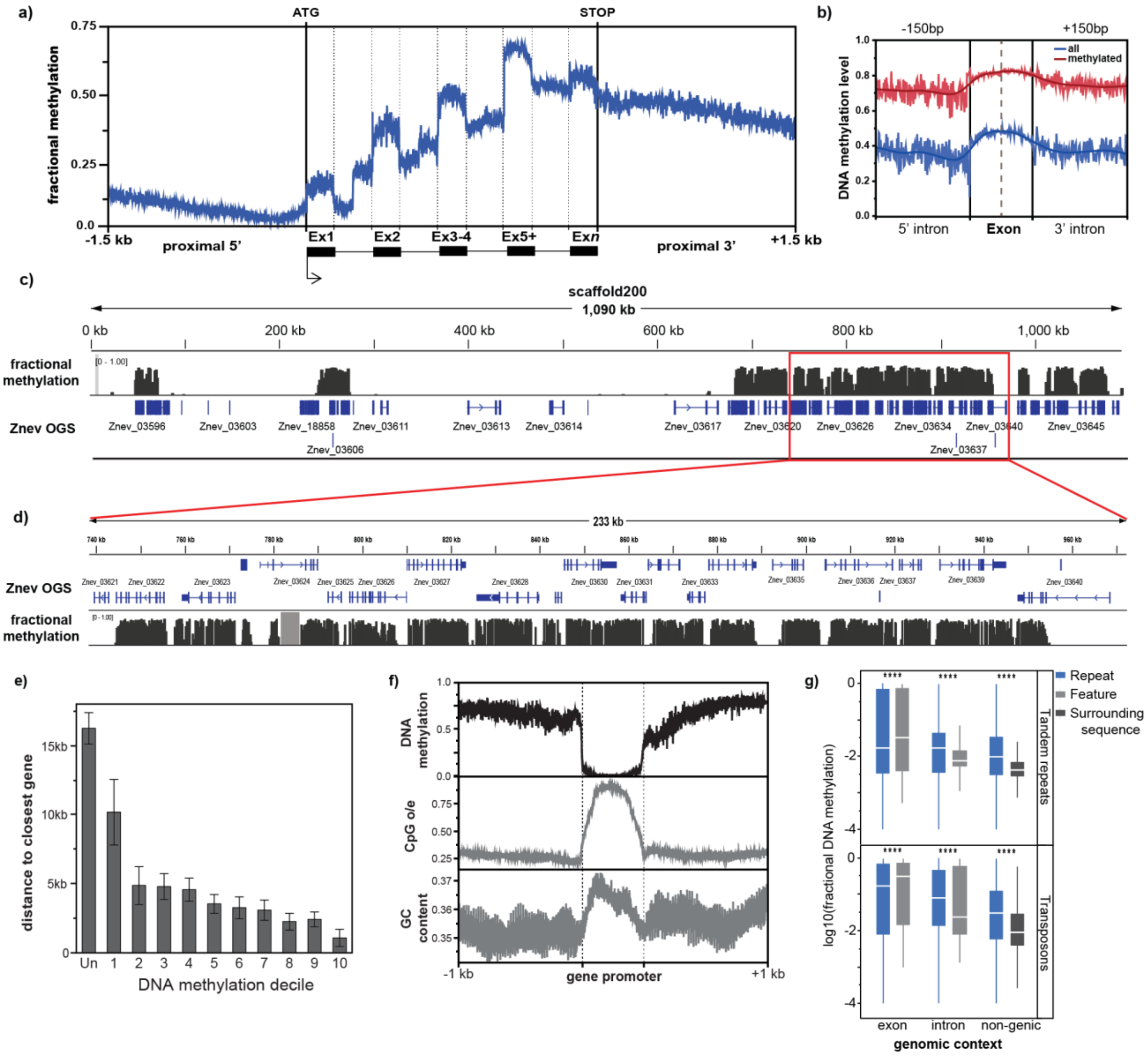


Figure 3.1. Genome wide DNA methylation patterns in *Z. nevadensis*. (a) Average methylation profile across multi-exon gene bodies of all *Z. nevadensis* genes with data, as well as (b) spatial profile of DNA methylation at exon-intron junctions for internal exons with matched data-containing up- and down-stream introns (150bp adjacent intronic sequence) across all exons (blue), as well as methylated exons only (red). (a) Genome browser snapshot of the entire scaffold 200, as well as (d) a 233 kb subset of scaffold 200, illustrating highly methylated gene-dense regions commonly seen in *Z. nevadensis*. (e) genomic distance to the nearest adjacent gene for 10 deciles of ascending DNA methylation level (1-10) as well as unmethylated genes (Un). (f) Average spatial plot of fractional DNA methylation, CpG o/e and GC content within and around unmethylated promoters of methylated genes. (g) Average DNA methylation level of repeats possessing 3 or more CpGs organized by repeats intersecting exons, introns, and non-genic repeats, as well as the average methylation level of the intersecting feature or adjacent non-repetitive genomic regions.

less variably expressed, but possess more alternative splicing events than methylated genes that do not differ. Furthermore, we find that caste- and sex-specific DMRs show significant enrichment for several *Drosophila* TFBS profiles, as well as multiple mature miRNAs, providing a potential mechanism by which differential methylation may impact the development of caste in termites.

Results

The termite DNA methylomes:

In order to examine DNA methylation both globally, and as it relates to differences between distinct castes, we performed sodium bisulfite sequencing for individuals from two castes (workers and alates). We further took advantage of our termite system by sequencing individuals from both sexes for each caste equally. We paired this with stranded RNA sequencing from the same morphs (castes x sexes), in order to explore what transcriptome differences were associated with differences in DNA methylation.

We found that DNA methylation in our termite samples existed at considerable levels (Figure 3.1, Figure B.1). Over 12% of genomic CpGs and 58% of exonic CpGs were methylated (Table B.4). The average quantitative methylation fraction averaged across all methylated and unmethylated exonic CpGs was 44.1%, and over 70% of genes featured significant methylation targeted to one or more exons (77.6% of genes as exons or introns). Notably, unlike the patterns seen in holometabolous insects with functional DNA methylation, where DNA methylation is preferentially targeted near gene starts (Hunt, Glastad et al. 2013), we found considerable methylation throughout methylated *Z. nevadensis* genes, with DNA methylation increasing as one progresses from 5’ → 3’ within the gene body (Figure3. 1a,c, Figure B.2). We further observed that regions downstream of methylated genes also exhibited considerable methylation (Figure 3.1a,c,d, Figure B.2, Figure B.3). DNA methylation was targeted to both exons and introns in *Z. nevadensis*, with considerable methylation in introns (Figure 3.1a-b, Figure B.1). Indeed, while a higher proportion of CpGs were methylated in exons than in introns, because of their size, over 2-fold as many mCGs exist within introns compared to exons (Table B.4). Nevertheless, exons did seem to possess higher methylation, although this difference is low when limiting the analyzed exons to those that possess data for both up- and down-stream introns (Figure 3.1b).

We found that, in general, methylated and unmethylated genes exhibited functional enrichment relative to one another, similar to that seen in other insects (Table B.7). Specifically, methylated genes were most highly enriched for functional terms related to fundamental cellular processes (eg ATP binding, DNA repair, histone modification), while unmethylated genes were associated with terms associated with more developmentally- or temporally regulated genes linked to organismal development (e.g. odorant binding, development of primary sexual characteristics, Wnt signaling pathway). We found that among all methylated genes, those most highly methylated (top 3 DNA methylation deciles; 3,072 genes) showed functional enrichment of terms associated with more fundamental regulatory terms such as “chromatin modification”, “protein binding”, helicase activity, metabolic processes, and “regulation of gene expression”, while the most lowly methylated genes (bottom 3 DNA methylation deciles) were functionally enriched for more dynamic terms such as “signaling receptor activity”, transmembrane transport of various molecules, cell periphery, and circadian behavior (Table B.8).

Genome-wide DNA methylation patterns in *Z. nevadensis*

Throughout the *Z. nevadensis* genome, the majority of methylated CpGs exhibited high fractional methylation, with 87.5% of methylated CpGs possessing >50% methylation fraction (Figure B.1). Similarly, genes tended to fall into two distinct classes, where some genes were very lowly methylated or unmethylated and others were highly methylated (Figure 3.1, FIgure B.1). When we examined gene annotations of

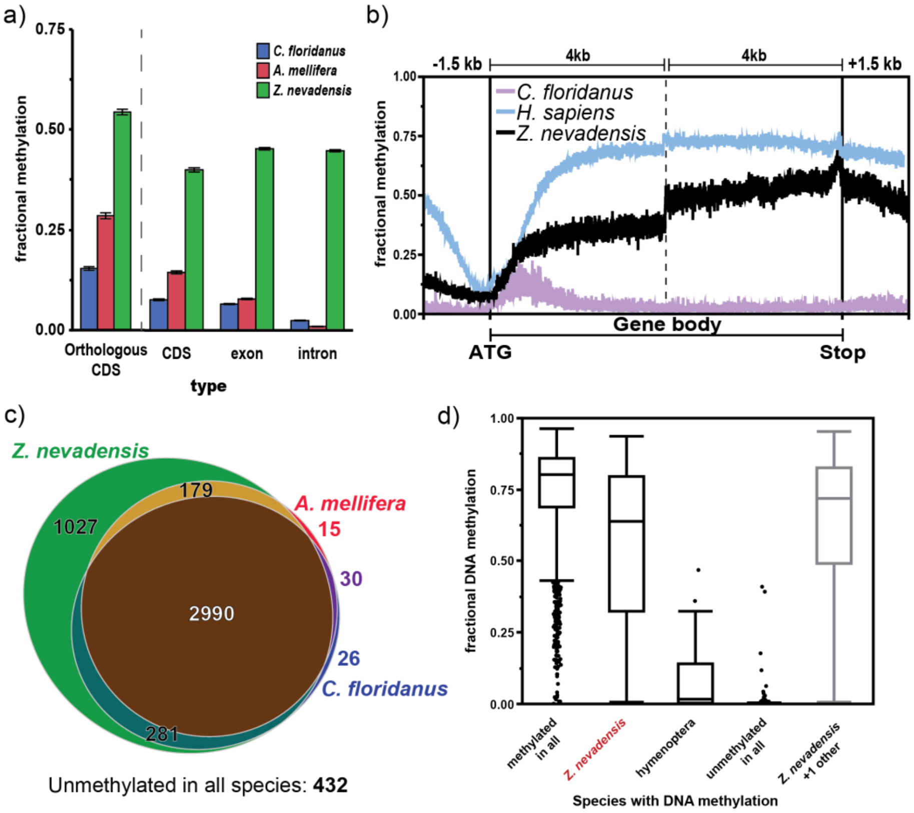


Figure 3.2. DNA methylation in *Z. nevadensis* exists at higher levels and is targeted to more genes than Hymenopteran social insects. (a) average fractional methylation for CDS (united exons) among conserved 1-to-1 orthologs between *C. floridanus*, *A.mellifera*, and *Z. nevadensis*, as well as all CDS, exons and introns for each species. (b) Average DNA methylation levels within the first and last 4kb of gene bodies for *Z. nevadensis* genes as well as for a representative hymenopteran (purple), basal invertebrate (red), and mammal (blue). (c) Venn diagram plot of methylation status among 4,931 conserved 1-to-1 orthologs showing large number of genes methylated only in *Z. nevadensis*. (d) Average fractional DNA methylation levels for genes classified based upon methylation status of the same orthologs and species from (c).

methylated and unmethlyated genes throughout the genome of *Z. nevadensis*, we observed that methylated genes tended to be clustered together, and were much more closely spaced than unmethylated genes (Figure 3.1d, Figure B.3). Indeed, throughout the *Z. nevadensis* genome we observed that these clusters of methylated genes result in large regions of highly-methylated CpGs, wherein regions up to hundereds of kilobases in size possess high levels of DNA methylation at the majority of CpGs (Figure 3.1c-d, Figure B.3). Within such methylation "blocks", the great majority of unmethylated CpGs correspond to gene promoters (Figure 3.1d,f). Furthermore, due to the increased evolutionary mutability of methylated cytosines, within such regions, CpG densities are depressed everywhere but within gene promoters (Figure 3.1f).

We found that several classes of repeats showed evidence of DNA methylation, but this analysis was complicated by the fact that many repeats fall within introns, which are often highly-methylated in *Z. nevadensis*. We thus examined genic (those intersecting exons or introns) and non-genic repeats separately. We found that among non-genic repeats, a low proportion of each repeat type showed at least some DNA methylation (although < 10% of non-genic tandem repeats), which was, overall, higher within non-genic transposable elements than in surrounding regions (Figure 3.1g, Figure B.5). Interestingly, for repeats intersecting genes, those intersecting exons appeared to be more lowly methylated than the exon they fell within (Wilcoxon rank-sum pvalue: <0.0001; Figure 3.1g). However, the majority of gene-intersecting repeats fell within introns, where repeats were more highly methylated than the containing intron (Fig 3.1g, Figure B.5). Despite this, we found that for both methylated exons and introns, those containing repeats were, in general, less methylated than those without (Figure B.5).

Orthology of methylation in *Z. nevadensis*:

Unlike in many insects examined to date, the *majority* of *Z. nevadensis* genes are methylated (74% methylated in at least one sample), and at higher levels (but see: Wang, Fang et al. 2014). In other social insect species that possess DNA methylation, approximately 1-2% of genomic CpGs are methylated (Glastad, Hunt et al. 2011). We found that over 12% of genomic CpGs and 58% of exonic CpGs were methylated (Table 3.1, Figure B.1). Utilizing DNA methylation data from the bee *A. mellifera* (Lyko, Foret et al. 2010) and ant *C. floridanus* ((Bonasio, Li et al. 2012)**;** methods) we contrasted DNA methylation in *Z. nevadensis* with DNA methylation in hymenopteran social insects, where approximately 35% of genes are methylated (Amel: 38.2% 4,946/12,961, Cflor: 35.7% 5,538/15,510). We found that genes that were methylated in the ant and bee tended to be methylated in *Z. nevadensis* (only 71 of 5,019 shared orthologs were methylated in ants and bees but not in *Z. nevadensis*, figure 2c, S7). In contrast, there were many genes methylated in *Z. nevadensis* that were unmethylated in *C. floridanus* and *A. mellifera* (1,027/5,019 shared orthologs; Figure 3.2c).

Genes methylated in *Z. nevadensis* but not ants or bees were highly enriched for terms relating to tissue- or temporal-specific gene expression (Table B.9). For example, we found that relative to genes methylated in all species, genes methylated only in *Z. nevadensis* showed greater than 10-fold enrichment for many terms, including “rhythmic process”, “sensory perception of chemical stimulus”, and “growth factor activity” (13.6, S14.8 and 22.16 fold enrichment, respectively). Despite the lineage specific methylation of these genes, their mean methylation level was considerable (median methylation fraction: 0.64, Figure 3.2d).

DNA methylation’s relationship with termite gene expression

We found that DNA methylation exhibited a similar relationship with gene expression as seen in other insects. Indeed, DNA methylation was positively associated with gene expression (rho: 0.348, R2: 0.179) and negatively correlated with expression variance between samples (rho: -0.356, R2: 0.155) and within samples (rho: -0.449, R2: 0.279; Figure 3.3b). However, in agreement with observed associations between gene body methylation and expression across diverse taxa

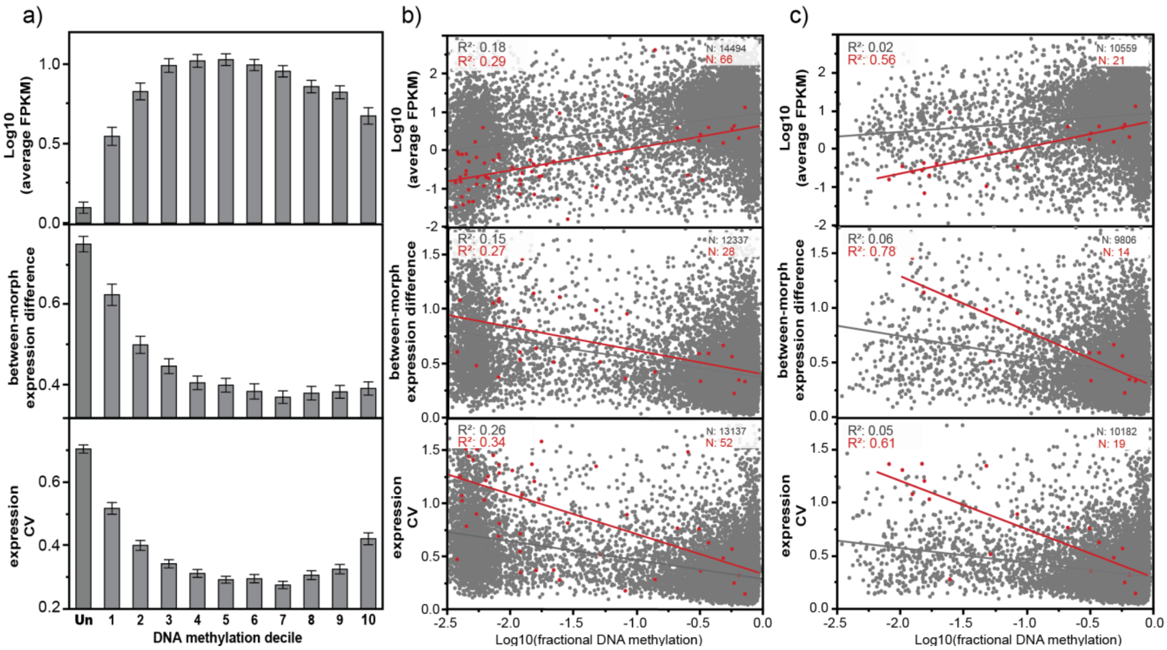


Figure 3.3. The relationship between DNA methylation and gene expression in *Z. nevadensis*. (a) Gene expression level, between-morph absolute expression difference, and within-morph replicate CV is presented for deciles of increasing DNA methylation (1-10) as well as unmethylated genes (Un). (b) Regression of the same variables as in (a) against a continuous measure of DNA methylation among all genes. (c) The same as in (b), but for methylated genes showing evidence of recent duplication in *Z. nevadensis* (red; from natcomms supp, N = 21), as well as for all other methylated genes (grey).

(Zemach, McDaniel et al. 2010; Glastad, Hunt et al. 2011), genes of intermediate methylation level were the most highly, and ubiquitously expressed (Figure 3.3a).

In order to test how DNA methylation is predicted by measures of gene expression in a combined framework, we performed regression analysis between DNA methylation level and our gene expression variables (level, CV, specificity), as well as intergenic distance (gene-gene distance), average exon count and general measures of gene conservation. We also leveraged the stranded nature of our RNA-seq protocol to produce a measure of anti-sense expression, incorporating this into our combined model as well. We found the strongest regressors in the combined model framework were expression CV and between-sample expression difference, which were both strongly negatively associated with genic DNA methylation level (Figure B.6). Interestingly, we found that when considered in this combined framework, the level of antisense gene expression was more strongly associated with DNA methylation level than gene expression level from the sense strand of a given gene model (Figure B.6). This suggests DNA methylation's association with gene expression level and variation may be driven, at least in part, by an interaction between DNA methylation and intragenic antisense transcription (Tufarelli, Stanley et al. 2003).

We next examined whether DNA methylation's relationship with gene expression among genes indicated as recently-duplicated (Terrapon, Li et al. 2014) differed from that observed globally among genes. We found that, among recently-duplicated genes with at least one methylated copy, the association between gene expression/breadth and DNA methylation level was much stronger than observed across all genes (Figure 3.3b). Furthermore, when considering only methylated genes, the relationship between DNA methylation level and gene expression variables was even stronger among recently duplicated genes, despite the strong reduction in association among all methylated genes (Figure 3.3c).

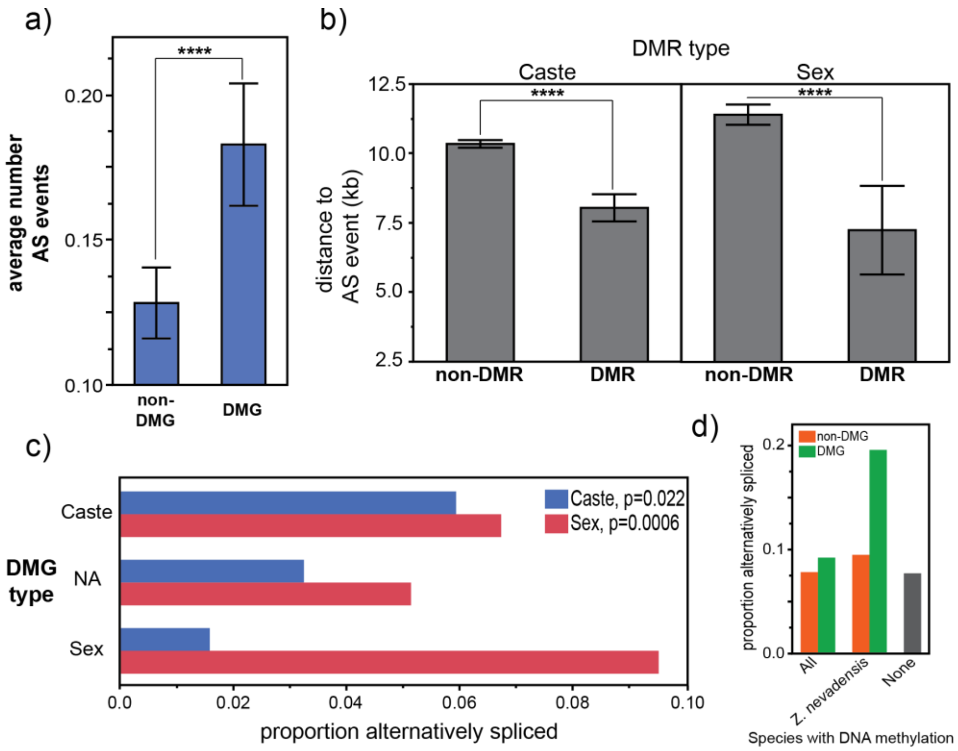


Figure 3.4. Differentially methylated genes show higher levels of alternative splicing. (a) Average number of alternative splice events observed among differentially methylated genes (DMG), as well as methylated genes that do not differ (non-DMG). (b) For both caste (left) and sex (right) DMRs, the distance to the nearest significantly differentially spliced exon is presented for DMRs and non-differing methylated regions (non-DMR). (c) for both caste- and sex-differentially methylated genes, the proportion of genes showing alternative splicing between castes (blue) and sexes (red) is given (p-values from fishers exact test). (d) The proportion of DMGs and non-DMGs featuring an alternative splicing event among genes methylated in all species, as well as those methylated only in *Z. nevadensis* (as in Fig2 c). Data is also presented for genes unmethylated in all species.

Differential methylation is strongly associated with caste:

We next evaluated differences in DNA methylation between our libraries. Because we sampled male and female samples from both castes, we were able to test for differential methylation between castes and sexes, while controlling for the opposite. Overall, we found 2,749 genes (of 10,974 tested genes) exhibiting significant differential methylation between one or more tested region in our combined (caste x sex) test. Interestingly, of the genes exhibiting differential DNA methylation, we found very few genes differentially methylated between sexes (210 genes), with the vast majority of differentially methylated genes (DMR-containing genes) existing between reproductives and workers (2,615 genes; Table 3.1). Even more interestingly, we found that of these caste-DMR-containing genes, the great majority exhibited higher methylation in the alate caste relative to workers (Table 3.1). Thus, the great majority of significant differences in DNA methylation between our four castes/sexes are composed of genes exhibiting higher methylation in the alate caste. We observed no difference in the Bisulfite conversion efficiency between libraries/morphs, as tested with a spike-in unmethylated lambda genome, as well as assessment of methylation rate at non-genic non-CpG cytosines (Table B.3).

DMR-containing genes are functionally enriched for multiple functional categories related to development and plastic response (Table B.11). For example, multiple development-associated terms show >2 fold enrichment among DMR-containing genes (e.g. embryonic pattern specification, motor activity, regulation of Rho signal transduction; Table B.11), with several terms showing >5-fold enrichment among DMRs (double-stranded RNA binding, regulation of cell projection organization, and GTPase binding; Table B.11).

We observed that differentially methylated genes show a signal of increased expression in the morph of hypermethylation across all genes with data (Figure B.9), however this association is weak, and not significant in every instance. Furthermore, we

Table 3.1: Numbers of differentially methylated and differentially expressed genes between castes and sexes, broken down into the morph of hyper methylation/expression.

	Test	**totals**	Hyper-morph	
			A	W
DMG	caste	**2,611**	2,515	96
			F	M
	sex	**209**	114	95
			A	W
DEG	caste	**1,094**	599	495
			F	M
	sex	**834**	792	42

found that overall, differentially methylated genes are significantly less variably expressed between replicates of the same morph, and show less *absolute* expression difference between morphs when compared to methylated genes that do not differ (Figure B.10). We further observe that, while methylated genes are in general much more conserved across insects than unmethylated genes (Figure B.10), differentially methylated genes are even more likely to be conserved across insects, and are less likely to be duplicated (both in general, and relative to average insect-wide copy number) when compared to methylated genes that do not differ (Figure B.19).

Notably, differentially methylated genes were more likely to feature at least one alternatively spliced exon, and the proportion of caste- or sex-specific alternatively spliced genes was highest for genes containing caste- or sex-specific DMRs, respectively (Figure 3.4a,c). We further found that within alternatively spliced genes also featuring at least one significant DMR, DMRs were located significantly closer to alternatively spliced exons than nonDMRs (Figure 3.4b). Finally, when we compared genes classified by their methylation status in multiple species (from above section), we found that genes methylated only in *Z. nevadensis* showed higher overall levels of

Table 3.2: Multiple TFBS motifs are enriched within DMRs. Presented are all TFBSs tested which possessed a putative ortholog within *Z. nevadensis,* and were significantly enriched within DMC-centered sequences when compared to control sequences using two methods. Also presented are the results of testing for enrichment of the given TFBS within caste- or sex-differing DMRs while using the alternative as control sequences (caste vs sex comparison), demonstrating that the majority of DMC-associated TFBS motifs are enriched within only one type of DMC relative to the other.

DMC type		TF	pvalue	caste vs sex comparison	Znev ortholog ID	Descriptive name
	Caste	Eip74EF	7.59E-06	**caste**	Znev_00833	Ecdysone-induced protein 74EF
		fkh	1.00E-05	**caste**	Znev_13477	fork head
		Ubx	3.93E-05	**caste**	Znev_15380	ultrabithorax
		bab1	7.15E-05	**caste**	Znev_03179	bric a brac
		en	5.28E-03	ns	Znev_15553	engrailed
	Sex	z	4.55E-02	**sex**	Znev_02821	zeste
		nub	8.55E-03	ns	Znev_14256	nubbin

ASing, when compared to genes methylated in all species. Furthermore, among such genes methylated only in *Z. nevadensis*, DMGs showed almost 2 fold more ASing than non-DMGs (Figure 3.4d).

Multiple TF and miRNA profiles are enriched surrounding Differentially methylated cytosines (DMCs):

Within gene bodies, DNA methylation has been proposed to have several functions, including dampening of spurious intragenic transcription. One way this may occur is through DNA methylation's ability to alter binding of TFs, either directly through methylated CpGs altering TF binding affinity at CpG-containing TF binding motifs, or through an alteration of nucleosome positioning at or nearby a TF binding site (Shenker and Flanagan 2012). Indeed, these same mechanisms are thought to underlie the observed strong negative association between promoter methylation and expression level of the associated locus (Suzuki and Bird 2008), as methylated promoters are associated with a less accessible chromatin state and an inability to initiate transcription (Deaton and Bird 2011; Jones 2012). Furthermore, intragenic DNA methylation in mammals has also been connected to nucleosome occupancy, TF binding, and intragenic transcriptional initiation (Lorincz, Dickerson et al. 2004; Maunakea, Nagarajan et al. 2010; Jones 2012; Shenker and Flanagan 2012).

Given the fact that DNA methylation within termites exists within introns and downstream of genes (as do a considerable number of DMRs: Table B.6), we sought to evaluate if differential methylation in our termite samples showed significant over-representation of transcription factor binding motifs. We first performed statistical tests examining the relative enrichment of existing *D. melanogaster* motif profiles (idmmpmm and flyreg profiles (Kulakovskiy and Makeev 2009)) within sequences centered on confidently differentially methylated cytosines (DMCs), relative to nearby sequences not showing significant differential methylation. We found that DMCs exhibited significant enrichment for multiple TF motifs taken from *Drosophila*, including several key

Table 3.3: DMRs are enriched for miRNA-similar sequences. Top 10 miRNA sequences showing overrepresentation among caste- and sex-biased DMRs, as determined by AME (FDR). Also provided are the number of significant hits (FIMO) for the given miRNA among caste- and sex-biased DMRs (+ hits) as well control sequences (non-significant tested regions within 1kb that do no overlap significant regions) – N *positive set,* caste: 5,786, sex: 1,364; N *negative set,* caste: 10,871, sex: 2,513. For both caste- and sex- biased DMRs, the number of positive and negative set sequences featuring a significant hit to *any* miRNA are also given (All miRNAs).

	miRNA	positive set hits	negative set hits	fold enrichment	AME FDR
Caste-biased DMCs	Zne-mir-34-3p	55	0	103.337	8.97E-13
	Zne-mir-263a-3p	24	0	45.092	1.11E-10
	Zne-mir-6012-5p	267	16	31.353	3.03E-15
	Zne-mir-2a-3-5p	10	0	18.788	1.39E-11
	Zne-mir-125-5p	10	0	18.788	1.54E-02
	Zne-mir-2796-3p	5	0	9.394	4.38E-02
	Zne-mir-279c-5p	4	0	7.515	5.97E-10
	Zne-mir-3049-5p	4	0	7.515	4.41E-06
	Zne-mir-87-1-3p	14	4	6.576	1.43E-05
	Zne-mir-981-5p	3	0	5.637	1.57E-12
	all miRNAs	556	136	7.681	
Sex-biased DMCs	Zne-mir-34-3p	19	0	35.005	4.68E-04
	Zne-mir-275-3p	9	0	16.581	4.13E-04
	Zne-mir-998-5p	7	0	12.897	6.44E-03
	Zne-mir-750-5p	6	0	11.054	3.80E-02
	Zne-bantam-5p	14	4	6.448	3.72E-02
	Zne-mir-6012-3p	3	0	5.527	1.58E-02
	Zne-mir-278-5p	3	0	5.527	1.79E-02
	Zne-mir-981-5p	2	0	3.685	1.09E-05
	Zne-mir-279c-5p	2	0	3.685	1.67E-02
	Zne-mir-184-5p	1	0	1.842	3.45E-05
	all miRNAs	222	110	3.718	

developmental TFs (with existing orthologs in *Z. nevadensis*). While the majority of these were enriched only among caste-specific DMCs, we observed two to be enriched surrounding sex-specific DMCs, with one (*zeste*) showing enrichment only for sex-specific DMCs, as well as relative towhen using caste-specific DMCs as control regions (Table 3.2).

We next we performed *de novo* motif identification (MEME) to indentify sequence motifs enriched among DMRs. We observed that the majority of most-significant identified motifs were very long (>15bp), and many exhibited considerable similarity to known *D. melanogaster* miRNAs (Table B.15). Because of this, we utilized putative mature miRNA sequences from *Z. nevadensis* as determined from the annotation of the genome (Terrapon, Li et al. 2014), to examine the enrichment of miRNAs as they exist in our focal species. We found multiple miRNA-like sequence motifs that were significantly more likely to be found within DMRs relative to control sequences (Table 3.3). Indeed, approximately 10% of DMCs we evaluated contained at least one significant hit to the profile of a mature miRNA (>6 fold enrichment of any miRNA among all caste-differing DMCs), with several specific miRNAs showing very strong overrepresentation among DMCs (Table 3.3). For example, approximately 4.6% of caste-biased DMC sequences (267/5,786) featured a significant hit to the miRNA zne-mir-6012, while only 0.15% of nearby nonDMC sequences featured such a hit (16/10,871).

For the great majority of significantly-DMC-associated miRNAs only one of the two mature miRNAs, produced from the associated miRNA hairpin, showed significant overrepresentation among DMCs. For example, for zne-mir-6012 which had a significant hit to 267 DMCs, *all of these* were to the 5-prime mature miRNA, with the 3-prime miRNA showing no hits within either DMC or control sequences (Table B.16). This suggests the miRNA profiles we observed enriched among DMCs represent a functional association, as for the great majority of studied miRNAs, only one of the two mature miRNAs produced from each miRNA precursor hairpin (sense- and anti-) is usually functional, despite being highly similar in sequence to one another.

Because miRNAs have classically been most strongly implicated in post-transcriptional silencing, through binding to 3'UTRs of mRNAs, we sought to examine the genic distribution of our miRNA-hitting DMCs, expecting many to fall downstream of genes within the putative 3'UTR. Interestingly, we found that the majority of DMCs that featured at least one significant miRNA profile hit fell within exons or introns (Table B.17; ~86%), with only ~8.5% falling within 2kb downstream of a given gene model. Thus, the majority of DMRs that feature significant similarity to miRNAs fall within gene bodies, and not 3'UTRs.

Discussion

We present the first survey of genome-wide DNA methylation in termites, and examine how it differs among both sexes of reproductive and worker castes. Because DNA methylation has been linked to the formation of caste in hymenopteran sociali nsects (Kucharski, Maleszka et al. 2008; Lyko, Foret et al. 2010; Bonasio, Li et al. 2012; Herb, Wolschin et al. 2012; Glastad, Hunt et al. 2014), and previous evidence suggested its existence in termites (Glastad, Hunt et al. 2013; Terrapon, Li et al. 2014), we sought to evaluate how this epigenetic mark relates to alternative phenotype in this developmentally-distinct, highly-diverged eusocial insect.

We report three major findings: (i) levels of DNA methylation exist at much higher rates than found in the majority of other insects examined to date, (ii) many significant differences in methylation exist between caste but few between sexes, (iii) caste and sex specific differentially methylated genes show higher levels of alternative splicing, and (iv) within genes differentially methylated regions are enriched for multiple regulatory motifs.

While DNA methylation was predominantly targeted to gene bodies as in other insects explored to date, in *Z. nevadensis* DNA methylation is targeted to more genes and exists at higher levels than seen in holometabolous insects (Figure 2 ; (Lyko, Foret et al. 2010; Bonasio, Li et al. 2012; Hunt, Glastad et al. 2013)). Indeed, we observe that methylated genes are often clustered together within the genome. Because DNA methylation targets the majority of the gene body (exons+introns), within such regions the majority of unmethylated CpGs correspond to gene promoters, resulting in the emergence of CpG-enriched promoters, surrounding by CpG-depleted, highly-methylated regions (Figure 3.1). Such promoters bear striking resemblance to vertebrate CpG islands. Thus, it is possible that DNA methylation in termites provides a glimpse at an ancestral state in the evolution of vertebrate CpG islands.

Our results further suggest DNA methylation targeting in *Z. nevadensis* is expanded relative to Hymenopteran insects. Indeed, the strong functional enrichment of genes methylated only in *Z. nevadensis* suggest that the expansion of DNA methylation in *Z. nevadensis* relative to holometabolous insects is associated with genes of specific function, with more developmentally or temporally-regulated expression than genes methylated in both Hymenoptera and Isoptera. DNA methylation patterning across *Z. nevadensis* gene bodies is similar to that seen in the basal invertebrate *C. intestinalis* (Zemach, McDaniel et al. 2010). Furthermore, the phylogenetic distribution of well-characterized DNA methylomes is highly biased to holometabolous insects (Glastad, Hunt et al. 2011), and preliminary evidence in several hemimetabolous insects suggest that DNA methylation may exist at higher levels in hemimetabolous insects than in holometabolous (Hunt, Brisson et al. 2010; Glastad, Hunt et al. 2013; Hunt, Glastad et al. 2013; Terrapon, Li et al. 2014; Wang, Fang et al. 2014). Thus, it is possible that the higher levels of methylation seen in termites are reflective of an ancestral loss of DNA methylation in the other species compared here, and not an expansion of methylation targeting in *Z. nevadensis*.

DNA methylation exhibited similar general relationships with gene expression level and variability (between and within samples) as seen in other insects with functional DNA methylation (Glastad, Hunt et al. 2011), and was most strongly negatively associated with expression noise when controlling for other factors. Notably however, we find the relationship between DNA methylation and expression level/breadth to be much stronger among recently-duplicated genes than all genes. This may suggest DNA methylation plays a novel role in regulating the expression of recently-duplicated genes, as suggested in (Wang, Wheeler et al. 2013). Alternatively, DNA methylation may simply correlate better with gene expression within rapidly sub-functionalizing gene duplicates due to the duplicate's rapid loss of function.

DNA methylation has been experimentally linked to the determination of caste in *A. mellifera* (Kucharski, Maleszka et al. 2008; Herb, Wolschin et al. 2012), and differs significantly between castes in ants (Bonasio, Li et al. 2012; Glastad, Hunt et al. 2014). When we compared DNA methylation between males and females of reproductive and worker termite castes, we found that the majority of differences in DNA methylation existed between castes, with far fewer existing between sexes. Notably however, we found that differential methylation exhibited a cryptic relationship with transcriptomic differences between the same sample types. Specifically, both caste- and sex-specific differentially methylated genes were less variably expressed, but showed considerably higher levels of alternative splicing than nonDMGs. Differences in DNA methylation in bees and ants are also linked to differences in alternative splicing, and it is hypothesized this relationship may underlie DNA methylation's impact on caste determination (Lyko, Foret et al. 2010; Herb, Wolschin et al. 2012). Importantly, both in our study and those in other social insects, the relationship between alternative splicing and differential methylation is somewhat weak, suggesting they may be linked through an indirect mechanism.

In mammals, gene body DNA methylation differences are also linked to differential splicing through several processes (Shukla, Kavak et al. 2011; Yearim, Gelfman et al. 2015). In most cases DNA methylation differences impact splicing through an alteration of TF binding. This can happen either directly through methylation's impact on TF binding at our around differentially methylated cytosines (DMCs) (Shukla, Kavak et al. 2011; Wang, Maurano et al. 2012), or by altering local chromatin, resulting in changes to DNA accessibility (Yearim, Gelfman et al. 2015). We found that multiple TFBS or TFBS-like motifs were enriched within or around termite DMCs, suggests that differential methylation is also linked to alterations in TF binding in termites. We further find that many of these binding sites exhibit enrichment in only one of either caste- or sex-specific DMC sequences. Notably, many of DMC-enriched TFs are associated with developmental processes in *D. melanogaster*.

Emerging evidence in model systems suggests that miRNAs can also impact the epigenome as well as the process of transcription (Li, Okino et al. 2006; WEINBERG, VILLENEUVE et al. 2006; Tan, Zhang et al. 2009; Wedeles, Wu et al. 2013), and several important components of the RNAi pathway have been shown to associate with chromatin in *Drosophila* in an smRNA-guided manner (particularly, euchromatin; (Cernilogar, Onorati et al. 2011)). Our finding that multiple mature miRNA or miRNA-like sequences exist surrounding DMCs is particularly intriguing, and suggests DNA methylation may also play a role in altering the binding of regulatory RNAs. Typically, miRNA genes produce two complimentary mature miRNA templates; however for the majority of miRNAs only one of these two templates is utilized by the components of the RNAi pathway. Thus, the fact that most DMC-associated miRNAs are only enriched for one of each pair of mature miRNAs supports the functional role of this association.

It is tempting to speculate that the preferential enrichment of specific motifs surrounding caste- and sex-DMCs reflects a major functional role for differential methylation in the phenotype-specific alteration of regulatory binding, as has been seen in model systems (Shukla, Kavak et al. 2011; Huh, Zeng et al. 2013). In support of this, many of the TF and miRNA motifs we observed to be significantly associated with termite DMC sequences have developmental functions in *D. melanogaster*. For example, two of the top caste DMC-associated TFs as well as one miRNA (Eip74EF, forkhead and miR-125) are directly associated with the regulation of ecdysone, an important molting regulatory hormone in *D. melanogaster* (Yamanaka, Rewitz et al. 2013). Ecdysone is also implicated in the regulation of caste in both hymenopteran social insects as well as termites (Terrapon, Li et al. 2014; Lavine, Gotoh et al. 2015), and shows biased expression between worker and alate termite castes). Nevertheless, a great many differentially methylated genes are not differentially expressed, and overall, differentially methylated genes actually show less expression difference between castes and sexes. Thus, exactly how differential methylation of these putative binding sites impacts phenotype is unclear.

Because this is the first single-base resolution genome-wide study of DNA methylation done in any termite (or any hemimetabolous insect, for that matter), follow-up studies will be necessary to better characterize the termite transcriptome and epigenome, as well as to better evaluate these findings as they relate to the wider gamut of termite castes. Nevertheless, our results add important insight into DNA methlyation's role in insect caste determination, as well as illustrate termites as an excellent model for future molecular studies of epigenetic underwriting of insect caste. Furthermore, our results highlight the general utility of termites as a developmental and evolutionary contrast to hymenopteran eusocial insects.

Materials and Methods

Samples and nucleic acid extractions

DNA and RNA were extracted from termite samples using standard DNA and cDNA extraction protocols. All samples were taken from a single colony, and contaminating gut material was removed prior to nucleic acid extraction. Bisulfite converted gDNA and stranded RNA were sequenced using the Illumina Trueseq platform

Read preprocessing and mapping

Raw RNA-seq and BS-seq reads were trimmed for quality and adapter contamination using Trimmomatic (Bolger, Lohse et al. 2014).

RNA-sequencing

Tophat2 (Trapnell, Pachter et al. 2009) was used to map strand-specific RNA-seq reads to the Znev genome (v1.0; (Terrapon, Li et al. 2014)). FPKM (fragments per kilobase of exon per million fragments mapped) produced by Cuffnorm was used to quantify expression levels at the level of the gene. Read counts for each locus were also established using the htseq-count script of the DESeq2 package (Love, Huber et al. 2014), and utilized for differential expression testing.

Cufflinks was also used to generate library-specific transcriptome annotations, which were then merged using cuffmerge. This merged cufflinks annotation was then resolved with the OGS annotations, and any multi—exon cufflinks transcript that did not overlap an OGS gene model were kept.

Bisulfite sequencing

Bisulfite-converted reads were mapped to the Znev genome using Bismark (v0.14.4; (Krueger and Andrews 2011), followed by duplicate removal. Reads were then used to infer methylation levels of cytosines genome-wide, using a binomial test, incorporating deamination rate (from an unmethylated control) as the probability of success, and assigned a significance value to each CpG site related to the number of unconverted reads (putatively methylated Cs) as they compare to the expected number from our lambda control. Resulting *P*-values were then adjusted for multiple testing (Benjamini and Hochberg 1995). Only sites with false discovery rate (FDR) corrected binomial *P* values < 0.01 featuring > 3 reads were considered “methylated”. Fractional methylation values were calculated, as described previously (Hunt, Glastad et al. 2013;

Glastad, Hunt et al. 2014), for each CpG site or for each genomic feature (exons and introns).

For the three other insect species compared here (*A. mellifera*, *C. floridanus*, *H. saltator*), trimmed sequencing reads were mapped to the respective genomes using Bismark. CpG DNA methylation was then quantified, and associated with features using the same methods as for *Z. nevadensis*. Orthologous relationships were then established between genes using orthodb (Waterhouse, Zdobnov et al. 2011) relationships for all genes with 0-1 copy in any species.

Differential methylation

Methylsig (Park, Figueroa et al. 2014) was used to assess differential methylation between samples. For both caste and sex, we assessed whether DNA methylation significantly differed using 200bp windows. We required at least 3 replicates of each caste (or sex) possess > 4 reads at tested CpGs/windows and be methylated in at least half of the samples, allowing for a total of 860,340 CpGs (among 175,410 windows) with sufficient coverage and methylation status.

We also performed the above analysis for all relevant caste and sex pairs (AF.WF+AM.WM and AF.AM+WF.WM, respectively), which we analyzed to produce a more conservative list of caste- and sex-associated DMR-containing genes that consistently, significantly differed between both pairs of a given comparison.

Differential expression testing

DESeq2 (Love, Huber et al. 2014) was used to assess differential expression at all annotated loci, using mapped read counts provided by htseq-count. Caste and sex were modeled as independent variables, allowing for the testing of each while condoling for the other, utilizing a likelihood ratio test. Only genes with at least 1 read in all samples were kept for testing. As for differential methylation, we also performed differential expression tests between each relevant caste or sex pair, which were further combined to establish genes consistently differentially expressed between both caste or sex pairs.

We also performed a separate analysis, incorporating two previously published soldier caste samples with equivalent replication across both sexes (3 male, 3 female;(Terrapon, Li et al. 2014)). For each of the three castes (alate, worker, soldier) we performed differential expression test comparing the focal caste to the remaining two (while controlling for sex), to identify putatively up- and down-regulated genes relative to the other two castes. We also performed a second test for differential expression between sexes (after controlling for caste), utilizing the samples from all three castes.

Differential exon usage

DEXSeq was used at assess differential expression of exons independent of differences in gene expression at the locus across all multi-exon genes after filtering out gene models lacking read coverage in >50% of samples.

Antisense transcription

In order to leverage the stranded nature of our RNA-seq protocol and roughly establish a measure for the level of antisense transcription, we first quantified the number of reads mapping to the sense and anti-sense direction of all gene models that do not overlap >50% of another gene mode. We then utilized a binomial-test method to identify significantly antisense transcribed genes (Balbin, Malik et al. 2015). That is, for each library we quantified the library-specific proportion of antisense read mapping, utilizing this value as the null binomial expectation. We then used a binomial test, using this library-specific null expectation (x2) to assess the probability that a given locus in a given sample exhibits antisense transcription no different from that observed across all loci. We then called each locus within each library as possessing significant antisense transcription based upon an FDR-corrected binomial pvalue < 0.05. Finally, we designated a locus as significantly expressed in antisense if >1/3rd of libraries exhibited significant antisense transcription.

We also produced a continuous metric of putative antisense transcription level for each caste+sex, by averaging the proportion of all reads that map to the antisense strand of a given gene across all three replicates of each caste+sex.

Orthology

We utilized Orthodb ortholog relationships for all ortholog groups with 1-to-1 representation in *A. mellifera*, *C. floridanus*, and *Z. nevadensis* (4,779 orthologs). In order to quantify large-scale patterns of gene gain and loss we also utilized orthodb (Waterhouse, Zdobnov et al. 2011) gene families from across all insect species represented on orthodb. For each gene family with representation in *Z. nevadensis*, calculated the average proportion of species with a member ortholog (large-scale conservation), average copy number of the given ortholog group across species (ancestral duplication rate), and the ratio of *Z. nevadensis* copy number to average cross-insect copy number (ancestral-normalized *Z. nevadensis* duplication rate). This allowed for the estimation of large-scale evolutionary patterns for each ortholog group, in lieu of alternative evolutionary metrics that are complicated by the absence of a closely related species with genome data.

MOTIF Detection

In order to detect motifs over-represented in differentially methylated regions, we first used the AME program (McLeay and Bailey 2010) to test for overrepresentation of *D. melanogaster* experimentally-established TF binding motifs (idmmpmm, 2009; flyreg (Bergman, Carlson et al. 2005)) and miRNAs (miRbase-dme (Kozomara and Griffiths-Jones 2014)).

For our test sets we extracted 150bp of genomic sequence surrounding confidently differentially methylated cytosines (FDR < 0.05, absolute methylation change > 20% between castes or sexes). For our control sets we did the same for all non-significantly differentially methylated cytosines falling within 1kb up- or down-stream of (but not overlapping) tested DMCs. This produced approximately 2x the number of control sequences for each test set.

In order to roughly quantify the fold-overrepresentation of a given miRNA within our test sequences (relative to control), we used the FIMO program (Grant, Bailey et al. 2011) to scan our test and control sequences for the given miRNA profile, then compared the test set size-normalized counts of significant hits (FDR < 0.1) between test and control sequences.

We further validated the results of our TFBS motif enrichment tests with the program Clover (Frith, Fu et al. 2004). For each TF binding profile we compared enrichment within our test sequences relative to both control sequences used above, as well as all methylated introns. We considered a TFBS motif confidently enriched within DMCs if both tests (AME and clover) showed the TFBS significantly enriched. Finally, for all TFBSs significantly enriched within DMCs, we further evaluated whether the given TFBS was enriched within caste- or sex-specific DMC test sequences relative to the alternative's test sequences (eg for caste DMC-surrounding sequences we used sex DMC-surrounding sequences as a control), in an effort to isolate highly-confident TFBSs enriched within phenotype-specific DMCs.

Acknowledgements

This work was supported by the U. S. National Science Foundation (DEB-0640690), the Georgia Tech-Elizabeth Smithgall Watts endowment, and the IUSSI Nutting Research Grant. We would also like to thank Brendan Hunt for invaluable comments throughout the drafting of this manuscript.

CHAPTER 4

DNA METHYLATION AND CHROMATIN ORGANIZATION IN INSECTS: INSIGHTS FROM THE ANT *CAMPONOTUS FLORIDANUS*[3]

Abstract

Epigenetic information regulates gene function and has important effects on development in eukaryotic organisms. DNA methylation, one such form of epigenetic information, has been implicated in the regulation of gene function in diverse metazoan taxa. In insects, DNA methylation has been shown to play a role in the regulation of gene expression and splicing. However, the functional basis for this role remains relatively poorly understood, and other epigenetic systems likely interact with DNA methylation to affect gene expression. We investigated associations between DNA methylation and histone modifications in the genome of the ant *Camponotus floridanus* in order to provide insight into how different epigenetic systems interact to affect gene function. We found that many histone modifications are strongly predictive of DNA methylation levels in genes, and that these epigenetic signals are more predictive of gene expression when considered together than when considered independently. We also found that peaks of DNA methylation are associated with the spatial organization of chromatin within active genes. Finally, we compared patterns of differential histone modification enrichment to patterns of differential DNA methylation to reveal that several histone modifications significantly covary with DNA methylation between *C. floridanus* phenotypes. As the first genomic comparison of DNA methylation to histone modifications within a single insect taxon, our investigation provides new insight into the regulatory significance of DNA methylation.

[3] Glastad, K. M., B. G. Hunt, et al. 2015. DNA Methylation and Chromatin Organization in Insects: Insights from the Ant Camponotus floridanus. Genome Biol. Evol. 7: 931-942.

Introduction

Most organisms are capable of developing different phenotypes in response to distinct environmental conditions. The molecular information regulating such developmental plasticity is often heritable through cell divisions, yet is not directly encoded by the genome. Transmission of such information is known as epigenetic inheritance (Berger, Kouzarides et al. 2009).

One of the most important forms of epigenetic information is the methylation of DNA. DNA methylation is present in all three domains of life (Klose and Bird 2006; Suzuki and Bird 2008; Glastad, Hunt et al. 2011), and has been linked to variation in gene regulation in mammals (Maunakea, Nagarajan et al. 2010; Shukla, Kavak et al. 2011), plants (Ecker and Davis 1986; Zilberman, Coleman-Derr et al. 2008; Zemach, McDaniel et al. 2010), and insects (Kucharski, Maleszka et al. 2008; Lyko, Foret et al. 2010; Li-Byarlay, Li et al. 2013). In mammals, DNA methylation has traditionally been associated with gene repression, particularly when localized to promoter regions (Bird and Wolffe 1999; Weber, Hellmann et al. 2007; Suzuki and Bird 2008). However, in mammals, plants, and even insects, methylation of DNA *within* gene bodies (exons + introns) is associated with actively expressed genes (Lyko, Foret et al. 2010; Maunakea, Nagarajan et al. 2010; Zemach, McDaniel et al. 2010; Glastad, Hunt et al. 2011; Shukla, Kavak et al. 2011). Notably, DNA methylation in insects is present at considerably lower levels than in plants or mammals, and is confined almost exclusively to gene bodies in holometabolous insects (Glastad, Hunt et al. 2011; Hunt, Glastad et al. 2013). Despite this, DNA methylation has been linked to the regulation of alternative developmental outcomes in social insects (Kucharski, Maleszka et al. 2008), potentially through its association with alternative splicing (Lyko, Foret et al. 2010; Shukla, Kavak et al. 2011; Flores, Wolschin et al. 2012; Herb, Wolschin et al. 2012; Li-Byarlay, Li et al. 2013).

DNA methylation acts in concert with other types of epigenetic information. For example, histone protein posttranslational modifications (hPTMs) also affect gene regulation and organismal development. Like DNA methylation, hPTMs have been found to mediate the binding affinities of protein complexes, such as those related to transcriptional and splicing machinery (Kolasinska-Zwierz, Down et al. 2009; Luco, Pan et al. 2010; Luco, Allo et al. 2011; Negre, Brown et al. 2011), as well as to control the local accessibility of chromatin (Henikoff 2008; Venkatesh, Smolle et al. 2012; Zentner and Henikoff 2013).

Until recently, genomic profiles of DNA methylation and hPTMs were not both available for a single insect species, making it difficult to gain insight into the integration of DNA methylation in the greater chromatin landscape. Nevertheless, comparative epigenomic studies revealed that patterns of DNA methylation grossly mirror patterns of several hPTMs across insect orders (Nanty, Carbajosa et al. 2011; Hunt, Glastad et al. 2013). These investigations suggest that DNA methylation acts in concert with hPTMs to affect gene regulation in insects, but the precise relationship between DNA methylation and hPTMs has yet to be explored. With the advent of genome-wide profiles of DNA methylation (Bonasio, Li et al. 2012) and hPTMs (Simola, Ye et al. 2013) for distinct castes of the Florida carpenter ant *Camponotus floridanus*, it is now possible to investigate how these two important classes of epigenetic modifications relate to one another at a fine spatial scale. Here, we interrogate the relationship between hPTMs and DNA methylation genome-wide in *C. floridanus* in order to better understand DNA methylation and its epigenomic context.

We find that hPTMs are highly predictive of DNA methylation in *C. floridanus*. In particular, a strong spatial relationship exists between highly methylated regions and patterns of hPTM enrichment within actively expressed genes. This relationship is further supported by an observed association, as assessed between social insect phenotypes, between differential DNA methylation and differential hPTM enrichment. Overall, these findings expand our understanding of the function of gene body methylation and how it interacts with other epigenetic information, such as that encoded by modifications to histone proteins.

Materials and Methods

Analysis of DNA methylation

DNA methylation level of genomic features

Genome wide, processed DNA methylation data for *Camponotus floridanus* were obtained from the Gene Expression Omnibus (GEO series: GSE31576, Bonasio, et al. 2012) for males, minor works and major workers (castes with associated ChIP-sequencing data). DNA methylation in animals is predominantly targeted to CpG dinucleotides (Yi and Goodisman 2009). Thus, fractional methylation levels were calculated as mCG/CG for each CpG, defined as the number of reads with methylated cytosines divided by the total number of reads mapped to the given CpG. FDR-corrected binomial p-values provided along with the CpG read data (Bonasio, et al. 2012 supplementary files deposited in GEO series: GSE31576) were used to assign a status of “methylated” or “unmethylated” to each CpG (FDR < 0.01). Only CpG sites with ≥ 4 reads were considered in analyses. Fractional methylation was calculated for specific genomic features (e.g., exons, introns) as the mean fractional methylation value of all CpGs within that feature. A feature was called as “methylated” if at least 3 CpGs within the feature were called as “methylated” according to the binomial test.

Determination of Highly Methylated Regions (HMRs) of the genome

We sought to detect Highly Methylated Regions (HMRs) of the genome, which we define as areas of high DNA methylation relative to much more lowly methylated regions directly up- and downstream of the HMR. HMRs were detected by identifying sharp transitions in DNA methylation levels using a sliding window method (length=250, step=50bp), wherein focal window DNA methylation level was compared to all windows within 500bp upstream (background). We determined that a focal window belonged to an HMR boundary if the focal window was greater than the background mean by a fractional DNA methylation level of at least 0.3, and if the difference between the focal window and the background mean exceeded 65% of the DNA methylation value of the focal window. Once established, an HMR boundary was extended to include all adjacent windows that exhibited a fractional methylation level greater than 50% of the level of the initial boundary window. This analysis was performed in both directions (5'to 3' and 3' to 5'), and resulting HMR boundaries were connected to form contiguous regions of high methylation, provided all windows either i) met the criteria for inclusion in both directional HMR boundaries, or ii) possessed a fractional methylation level ≥ 50% of the mean of both boundaries. Unpaired HMR boundaries were themselves called as HMRs provided they did not fall within 500bp of another HMR and possessed at least 4 methylated CpGs (according to the binomial test). Orientation was established by finding the closest gene (up to 2kb) to a given HMR and assigning that HMR its strandedness (Glastad, Hunt et al. 2011) – HMRs not falling within 2kb of a gene were not assigned a strand.

HMRs in the genome were then compared with gene annotations (Cflo_OGSv3.3) and assigned a status of "exon", "intron", "5'-upstream", or "NA" (not overlapping a genic future), as well as being called as "5'-proximal" (≤ 1500 bp from start codon) or "non-5'-proximal" (any other genomic region).

Determination of Differentially Methylated Regions (DMRs) of the genome between castes

We identified regions of the genome that were differentially methylated (DMRs) between the male and worker castes by examining 200bp windows (step = 100bp) between each pairwise comparison of castes (due to the very low number of DMRs (12) identified between minor and major worker castes, we only considered comparisons between males and workers). We modeled methylation levels for each genic feature as a function of two categorical variables: "caste" and "CpG position" using generalized linear models of the binomial family (GLM), implemented in the R statistical computing environment (R Development Core Team 2011). If caste contributed significantly (chi-square test of GLM terms, p-value < 0.01) to the methylation status of a window (after adjustment for multiple testing using the method of Benjamini and Hochberg 1995), the window was considered differentially methylated between castes (Lyko, Foret et al. 2010). Only CpG sites that were significantly methylated (after multiple test correction) in one or both castes and covered by ≥ 4 reads in both libraries were used in these comparisons. Moreover, only features with ≥ 3 CpG sites were considered in these analyses. Once regions were assigned as DMRs, each DMR was then called as "elevated" in the caste with higher fractional methylation level. Overlapping windows of the same differential methylation status (Caste1 > Caste2, Caste2 > Caste1, or not differentially methylated) were then combined.

Analysis of histone modifications

ChIP-seq read alignment and signal estimation

ChIP-sequencing data are the product of preferential enrichment of gDNA bound to a specific chromatin protein. For each hPTM, raw sequencing reads are processed followed by alignment to the reference genome of the organism in question. Once aligned, reads reflect quantitative levels of ChIP signal that can then be further normalized to a no-antibody (input) control to produce a base-wise measure of the enrichment of ChIP signal reads over the control library – reflective of protein binding or prevalence (Park 2009).

We analyzed the prevalence of hPTMs H3K4me1, H3K4me3, H3K9ac, H3K9me3, H3K27ac, H3K27me3, H3K36me3, as well as the protein RNA polymerase (pol) II, in males, minor workers and major workers (Simola, Ye et al. 2013). After quality and adaptor trimming (trimmomatic: (Bolger, Lohse et al. 2014)), raw sequencing reads (accession: SRX144014-SRX144044) were mapped to the *C. floridanus* genome (v3.0) with bowtie2 (Langmead, Trapnell et al. 2009) using the options "--sensitive -k 1 -

N 0”. MACS2 (Zhang, Liu et al. 2008) was then used to estimate the read enrichment relative to an input control (as well as bulk histone H3 profiles for histone modifications to histone H3) for each ChIP library after removal of any duplicate reads using samtools (Li, Handsaker et al. 2009). Unless otherwise noted, all general comparisons between DNA methylation and hPTMs employed DNA methylation and hPTM enrichment averaged across all 3 castes.

Determination of peaks of ChIP-enrichment

Regions of significant ChIP signal enrichment (ChIP enrichment “peaks”) in the genome were established using MACS2 (FDR < 0.01), which identifies regions significantly enriched with a given ChIP signal relative to control libraries. Such peaks indicate regions that are likely to be strongly bound by a given chromatin protein. We considered a feature (e.g., exon, intron) to be significantly bound with a given protein if >10% of its length was overlapped by a region of significant enrichment for that mark.

Determination of regions of differential ChIP enrichment between castes

Differentially bound regions (DBRs) were established using the program MAnorm (Shao, Zhang et al. 2012), which uses common peaks between two libraries (as called by MACS2) to rescale and normalize ChIP data between two treatments, then estimate significance, direction and magnitude of differential ChIP enrichment for all confident ChIP enrichment peaks. Candidate DBRs with an FDR corrected p-value of < 0.01 were called as differentially enriched between castes, and the direction of differential binding enrichment was determined from the MAnorm-produced normalized between-comparison ChIP enrichment M-value ($\log_2$ ratio).

Analysis of gene expression

We determined levels of expression for given genes by analyzing RNA-seq data from the three castes which also have DNA methylation and ChIP-seq data (male, minor worker, major worker)(Bonasio, Zhang et al. 2010). Raw RNA-seq reads (GSM563074, GSM921123, and GSM921122) were filtered and aligned to the *C. floridanus* genome (v3.3; (Bonasio, Zhang et al. 2010)) using Tophat (Trapnell, Pachter et al. 2009), with the options "-r 50 --mate-std-dev 11(/20) -i 60 --no-discordant --read-realign-edit-dist 0 --coverage-search --b2-sensitive" specified. Cufflinks (Roberts, Pimentel et al. 2011) was run with multi-read and fragment bias correction ("-u" and "-b" respectively), and upper quartile normalization was used. Assemblies across castes were merged using cuffmerge ("-s"). FPKM (fragments per kilobase of exon per million fragments mapped) produced by Cuffdiff was used to quantify expression levels at the level of the gene.

Combined analysis of DNA methylation, ChIP analysis, and gene expression

We investigated if the patterns of DNA methylation were correlated with the presence of chromatin proteins in *C. floridanus*. In order to do so, we used measures of mean fractional DNA methylation level and average normalized ChIP enrichment for each coding sequence (CDS) to perform linear regressions and Spearman's rank correlations between epigenetic marks with the JMP statistical software package (SAS Institute Inc.). For each hPTM we determined the correlation coefficients derived from its correlation with DNA methylation among all CpGs (allCpG), as well as among only those CpGs determined to have at least some significant DNA methylation (mCGs).

We next determined patterns of ChIP-seq enrichment relative to HMRs. ChIP-seq enrichment was calculated for each HMR, as well as for 0.5kb regions up- and downstream of each HMR in order to identify relationships between levels of DNA methylation and the presence of hPTMs. For analyses of ChIP enrichment profiles relative to HMR boundaries, continuous ChIP enrichment signal was averaged at each base up to 1kb up- and down-stream of HMR boundaries. Within HMRs, length-proportional bins were used to average between HMRs – allowing for differing HMR lengths.

We next investigated if there were relationships between DMRs and DBRs between *C. floridanus* castes. We first compared DMRs to DBRs genome-wide, in order to test whether DMRs are preferentially associated with DBRs. We tested for enrichment of DBRs among DMRs, relative to non-DMRs, using a Fisher's exact test. We then tested if the directionality of a DMR showed any significant association with the direction of differential ChIP enrichment at that locus. For each caste pair we assigned each DMR and DBR the caste which showed the highest pairwise DNA methylation or ChIP enrichment levels, respectively, and then determined if hypermethylation in a specific caste was associated with consistent increases or decreases in that caste's ChIP enrichment at the same locus.

Finally, we were interested in understanding if epigenetic factors, including hPTMs and DNA methylation, were jointly predictive of patterns of gene expression. In order to evaluate the contributions of DNA methylation to gene expression level, we performed multiple regression analyses between the epigenetic marks (methylation + hPTMs) and gene expression. We first performed regressions between gene expression and each mark independently. We then performed regression using all epigenetic marks in a multiple regression model. For single-term tests, each factor was regressed against gene expression (log2(FPKM+0.01)) and bias independently, then for the full test as a component of an additive model including all factors. This enabled a comparison of DNA methylation's contribution to gene expression when controlling for hPTM enrichment and vice versa. All variables were standardized (0-centered after normalization) before model fitting.

Results and Discussion

DNA methylation is strongly associated with active histone modifications

Recent studies in plants (Zilberman, Coleman-Derr et al. 2008; Zemach, McDaniel et al. 2010; Coleman-Derr and Zilberman 2012) and animals (Ooi, Qiu et al. 2007; Cedar and Bergman 2009; Shukla, Kavak et al. 2011) have demonstrated that epigenetic information encoded by DNA methylation and hPTMs may interact to affect gene function. We thus sought to evaluate the relationships between DNA methylation

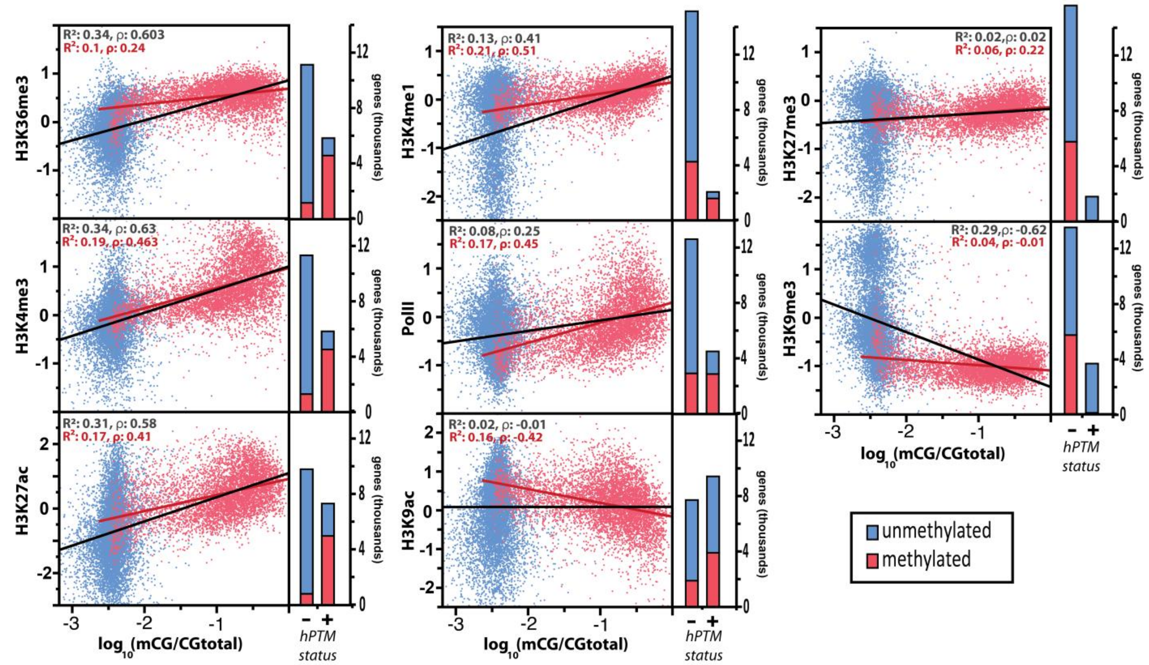


Figure 4.1. Histone modification enrichment as a function of DNA methylation levels for methylated (red) and unmethylated (blue) genes. Linear fits for all genes (black line) and methylated genes only (red line) are provided, along with their relevant R^2 values and Spearman's correlations, ρ. Bars represent the number of genes belonging to each class: those with significant histone posttranslational modification (hPTM) enrichment (+) and those without (-).

and hPTM enrichment in the *C. floridanus* genome, and thereby improve our understanding of insect gene regulation.

Each hPTM we investigated was significantly over- or under-represented among methylated genes (Figure 4.1, Table C.1). Consistent with previous comparative results (Nanty, Carbajosa et al. 2011; Hunt, Glastad et al. 2013), the hPTMs that are generally most strongly associated with actively expressed genes (H3K4me3, H3K27ac and H3K36me3;(Kharchenko, Alekseyenko et al. 2011)) were highly over-represented among methylated genes. H3K4me3, H3K27ac, and H3K36me3 were present among over 79% of methylated genes, with 95% of methylated genes featuring at least one of these hPTMs (Table C.1). Conversely, repressive hPTMs (H3K27me3 and H3K9me3;(Kharchenko, Alekseyenko et al. 2011)), which are generally associated with much less broadly expressed genes, were significantly and strongly under-represented among methylated genes, with less than 2% of methylated genes significantly enriched for either modification (Figure 4.1, Table C.1).

Similarly, when examining correlations between coding sequence DNA methylation *levels* and hPTM enrichment we found that the level of gene methylation was strongly positively associated with the quantitative level of ChIP enrichment for the active hPTMs H3K4me3, H3K27ac, H3K36me3, and H3K4me1, as well as for RNA polymerase II (RNA pol II) (mean rho: 0.53; Figure 4.1 and Table C.2). Conversely, the repressive hPTM H3K9me3 was strongly negatively correlated with coding sequence DNA methylation levels (rho = -0.62; Figure 4.1 and Table C.2). Thus, within insect genomes DNA methylation shows strong preferential targeting relative to most well-studied hPTMs, and is strongly biased to genes exhibiting active hPTMs. Consistent with this finding, hPTM levels explained 70% of the variance in coding sequence DNA methylation as inferred by the R^2 value generated by multiple regression (Figure C.1). We observed that many of the correlations between overall coding sequence methylation level and hPTM enrichment largely result from the fact that genes featuring

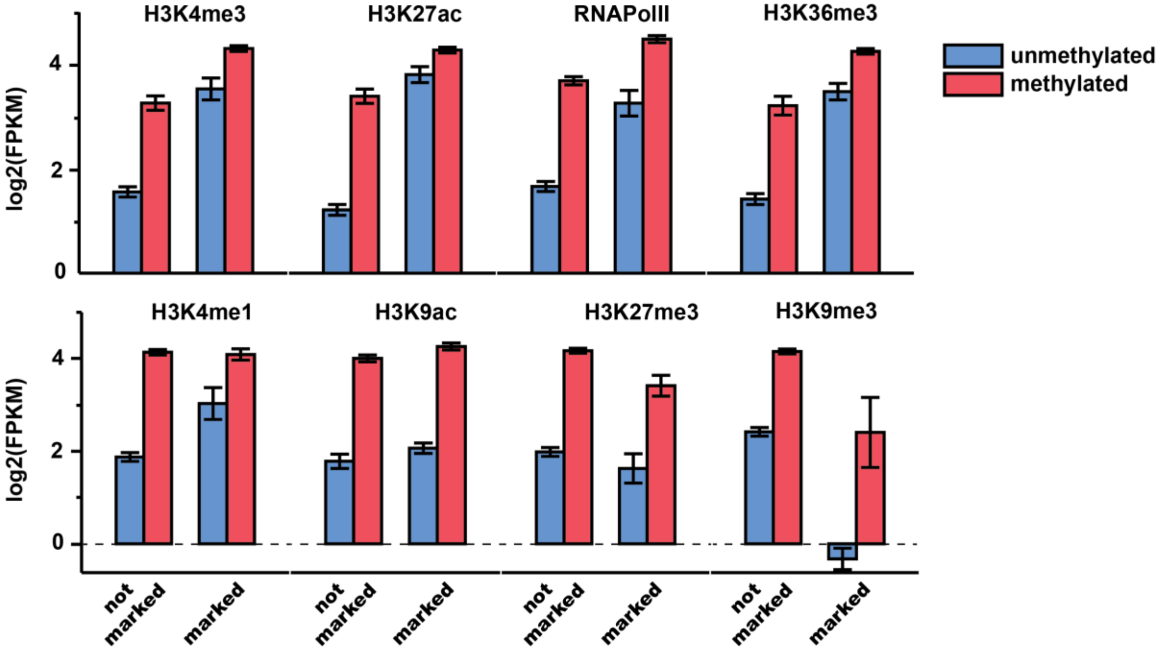


Figure 4.2. Methylated genes are more highly expressed, independent of hPTM status. Expression levels (log$_2$(FPKM)) of genes associated (marked) or not associated (not marked) with particular histone modifications. Methylated genes exhibit consistently higher expression relative to unmethylated genes, regardless of their associations with particular histone posttranslational modifications.

any DNA methylation were also those most likely to exhibit significant regions of enrichment or depletion of hPTMs (i.e., binary associations; Figure 4.1). Consequently, when limiting our analysis to only genes displaying significant levels of DNA methylation, we found that many correlations between DNA methylation and hPTM enrichment were substantially weakened (Figure 4.1). hPTMs associated with actively expressed gene TSSs (namely, H3K4me3 and H3K27ac) and RNA pol II, however, maintained relatively strong relationships with DNA methylation level among significantly methylated genes (Figure 4.1). Interestingly, despite being considered an "activating" mark and being significantly co-localized to methylated genes, the hPTM H3K9ac exhibited a considerable negative correlation with DNA methylation in this methylation-limited analysis. This may be due to DNA methylation's tendency to be most highly targeted to genes of intermediate expression, while H3K9ac is known to target very highly expressed genes. Moreover, a previous analysis found H3K9ac to be strongly preferentially targeted to high-CpG regions within promoters (Supplementary Figure 9 of Simola, et al. 2013), which are also the most consistently depleted of methylation.

Finally, though we observed strong relationships between DNA methylation and hPTMs at the gene level, we sought to evaluate the presence of direct spatial overlap between epigenetic marks within genes. We found that the observed relationships between DNA methylation and specific hPTMs remained largely intact when considering DNA methylation enrichment within regions of significant hPTM enrichment (Figure C.2) or within spatially-restricted windows downstream of the TSS (Figure C.3).

Overall, active hPTMs seem to be highly predictive of genic DNA methylation levels. That is, active hPTMs are (i) targeted to the same loci as DNA methylation, (ii) positively correlated with DNA methylation levels at these loci, and (iii) spatially enriched for DNA methylation within hPTM-marked regions. The hPTM most consistently and strongly associated with DNA methylation in our analyses was H3K4me3 (Figure 4.1, Tables C.1 and C.2).

DNA methylation and histone modifications bear similar, but non-redundant, associations with gene expression

We next sought to evaluate how DNA methylation and hPTMs were related to patterns of gene expression in the broader context of the other epigenetic information studied here. We compared gene expression levels between genes possessing at least one region significantly enriched for a given histone modification and/or DNA methylation in order to evaluate the redundancy of DNA methylation to individual hPTMs in explaining gene expression levels. We found that, among genes possessing at least one region significantly enriched for a given histone modification, those with DNA methylation exhibited consistently higher expression levels and consistently lower expression bias than those with the same modifications but no DNA methylation (Figures 4.2 and C.4).

We sought to further evaluate how epigenetic factors and their interactions related to gene expression in a combined framework using multiple regression analysis. We investigated if hPTMs and DNA methylation were predictive of gene expression level and gene expression bias among castes, as measured by RNA-seq. We first performed regressions between each epigenetic mark and gene expression separately. Not surprisingly, DNA methylation showed a significant positive association with gene expression when regressed singly (Table 4.1). Moreover, when incorporated into a full regression involving all epigenetic marks, DNA methylation still contributed significantly to the modeling of gene expression. This indicates that, even after accounting for the contribution of hPTMs, DNA methylation remains independently associated with gene expression (Table 4.1). Thus, though DNA methylation is highly correlated with active hPTMs, methylated genes were more highly and broadly expressed than unmethylated genes, even when controlling for hPTM status.

Table 4.1. Regression analysis for estimating gene expression level and bias from epigenetic marks. Coefficients for both single-term tests and full model are provided. All single-test model parameters significant at the P < 0.0001 level. P≥ 0.05, *; P< 0.05, **; P< 0.01, ***; P< 0.001, **** P< 0.0001. N = 15,165.

	Gene expression level			Gene expression bias		
Effect	R^2 (single term)	Coefficient (single term)	Coefficient (full model)	R^2 (single term)	Coefficient (single term)	Coefficient (full model)
DNA-methylation	0.279	1.875	0.424^{****}	0.165	-0.401	-0.170^{****}
H3K4me3	0.273	1.869	-0.128^{**}	0.151	-0.476	0.162^{****}
H3K4me1	0.222	1.684	0.238^{****}	0.086	-0.504	-0.067^{***}
H3K27me3	0.081	1.020	-0.537^{****}	0.002	0.021	0.161^{****}
H3K27ac	0.343	2.096	0.891^{****}	0.207	-0.567	-0.272^{****}
H3K36me3	0.344	2.097	1.382^{****}	0.205	-0.618	-0.373^{****}
H3K9me3	0.307	-1.983	-0.610^{****}	0.279	0.723	0.233^{****}
H3K9ac	0.082	1.022	-0.119^{**}	0.084	-0.390	-0.255^{****}
PolII	0.124	0.558	-0.256^{****}	1.22E-05	-0.010^{****}	0.164^{****}
R^2 adj. (full model)	**0.5086**			**0.4126**		

Histone modifications are strongly spatially organized relative to regions of DNA methylation in insect genomes

Up to this point, we have described associations between DNA methylation and hPTMs as summarized at the level of genes. These analyses provide important insight into the co-association of DNA methylation and hPTMs as it relates to patterns and levels of gene expression. However, such analyses are unable to provide insight into the precise localization of DNA methylation and hPTMs, let alone their interplay. Thus, we sought to evaluate levels and patterns of hPTM enrichment at a fine spatial scale relative to highly methylated regions. This facilitates an evaluation of hPTM enrichment within the spatial context of DNA methylation, but independent of other genomic annotations (gene features, etc). To accomplish this aim, we first developed an algorithm to establish regions of high fractional DNA methylation bordered by regions of much lower DNA methylation (see methods). This produced a set of 7382 Highly Methylated Regions (HMRs), which were subsequently analyzed for hPTM enrichment.

HMRs represented highly methylated regions in the otherwise-sparsely methylated *C. floridanus* genome, with an average fractional methylation level of 0.63, and almost 70% of individual highly methylated CpGs (CpGs with >0.5 fractional DNA methylation) fell within an HMR. Despite this, HMRs were only an average of 650.3bp (SD: 335.6bp) long, and while over 85% of genes with significant DNA methylation featured at least one HMR (4922/5785 methylated genes), HMRs only covered about 33% of the area of these genes. Thus, even within methylated genes, regions of high methylation are often limited to only a portion of the gene, most frequently at the 5’ end of these genes (Bonasio, Li et al. 2012; Hunt, Glastad et al. 2013). As expected, out of our 7382 HMRs, the great majority (6927; 93.8%) were located within or near genes, and only 22/7382 of such peaks did not fall within 2kb of a gene annotation or RNA-seq-based cufflinks annotation (Table C.3). Of these 22, only 14 showed no RNA-

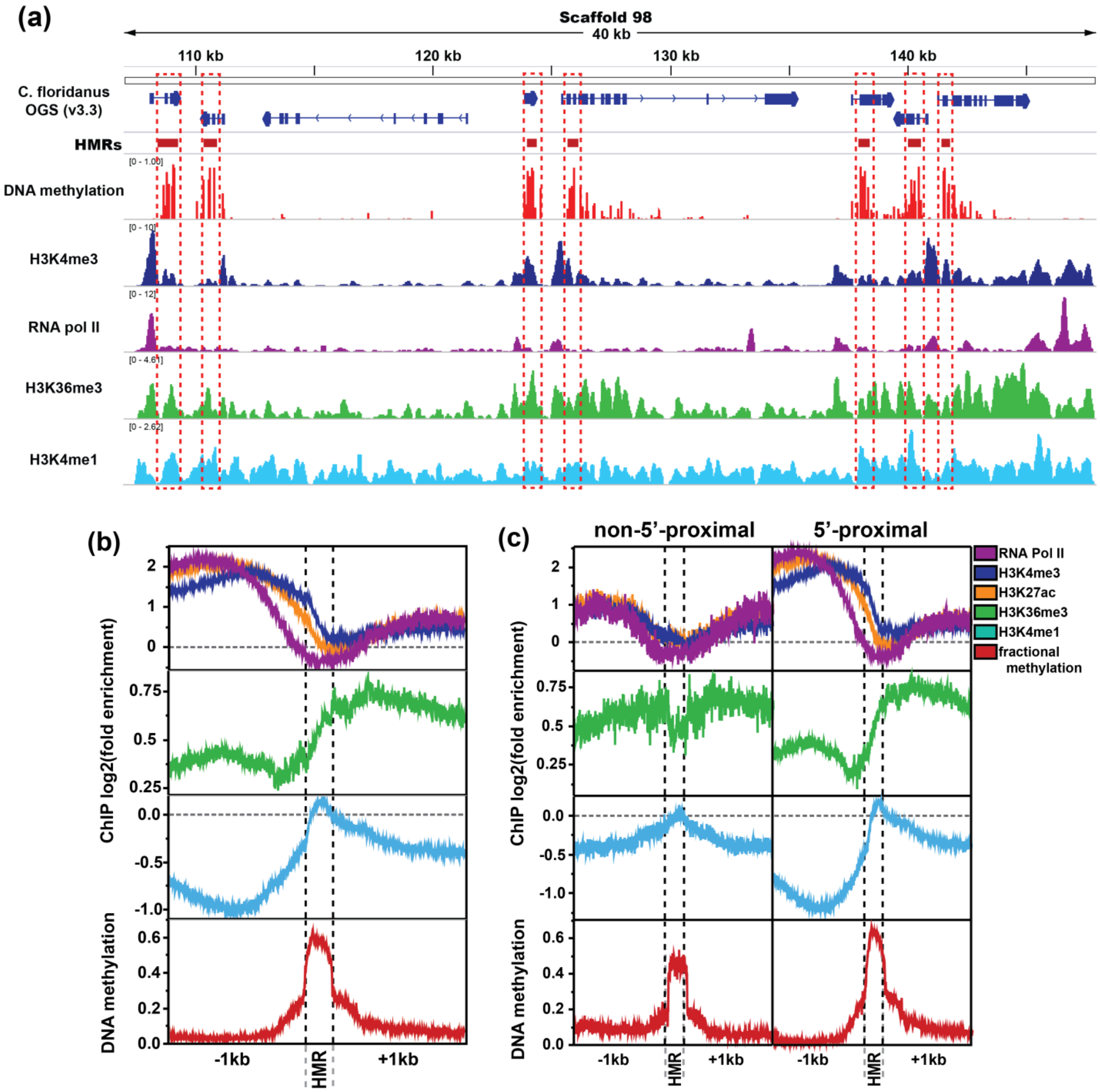


Figure 4.3. Active histone modification enrichment significantly differs between highly methylated regions (HMRs) and non-HMRs. (a) Example genome browser track showing stark spatial contrast between DNA methylation (HMRs) and promoter-proximal active chromatin (highlighted in red boxes). (b) Spatial relationship between DNA methylation and select histone posttranslational modifications or polymerases. (c) ChIP enrichment at HMRs separated by whether they fall within 1200bp of a gene start ("5'proximal") or not ("non-5'proximal"). Significance values represent results of Kruskal-Wallis tests comparing HMRs and 500bp regions in each direction of HMR boundaries. Repressive hPTMs showed little organization relative to HMRs, and were thus excluded.

sequencing coverage from the samples analyzed here. Thus, the overwhelming majority of HMRs are associated with expressed genes.

Studies of hPTMs in *C. floridanus* and other insects have revealed that many hPTMs, particularly those associated with actively transcribed genes, exhibit a strong spatial organization relative to the TSS of genes (Kharchenko, Alekseyenko et al. 2011; Simola, Ye et al. 2013). TSSs and surrounding proximal regions of active genes are marked with highly-accessible chromatin and enriched with the hPTM H3K4me3. In contrast, further-3’ regions of the same transcribed genes are marked with the hPTM H3K36me3, indicative of less-accessible regions of chromatin characterized by transcriptionally-elongating RNA pol II (Bannister and Kouzarides 2011; Kharchenko, Alekseyenko et al. 2011). Recent investigations have revealed that DNA methylation in *C. floridanus* and other holometabolous insects is preferentially targeted to the 5’ region of genes, immediately downstream of the TSS (Bonasio, Li et al. 2012; Hunt, Glastad et al. 2013). The common spatial organization of active hPTMs and DNA methylation relative to gene starts suggests a functional interdependence between DNA methylation and hPTMs within actively expressed insect genes.

Consistent with this idea, we found that HMRs exhibited significantly different levels of enrichment for most active hPTMs relative to regions directly up- and down-stream of HMRs (Figures 4.3, Figure C.5). More specifically, HMRs tend to lie between distinctive promoter- and gene body-associated hPTMs: TSS-associated active hPTMs, including H3K9ac, H3K4me3, H3K27ac, as well as RNA pol II, were enriched upstream of HMRs, while H3K36me3 was depleted upstream and enriched downstream of HMRs (Figures 4.3b-c,Figure C.5). For these active hPTMs we also found that the level of HMR methylation correlated positively with quantitative levels of ChIP enrichment within or nearby HMRs (Figure 4.4), indicating a strong quantitative link between hPTM enrichment and DNA methylation at a local level. Notably, we found that active TSS-

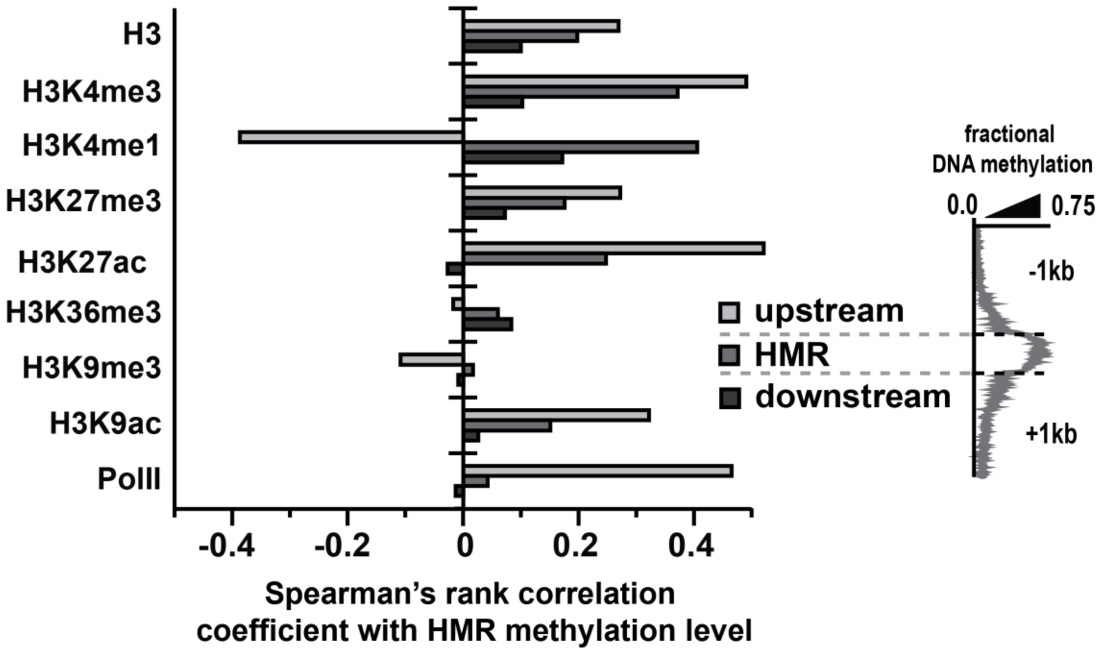


Figure 4.4. Distinct associations between DNA methylation and hPTMs upstream and downstream of highly-methylated regions (HMRs). Spearman's rank correlation coefficients between methylation level for HMRs and histone posttranslational modification enrichment within HMRs and 1kb up- and downstream of HMRs.

associated hPTMs were most strongly correlated with HMR methylation level directly upstream of the HMR, and not within the HMR itself (Figure 4.4).

The TSS-proximal boundary between H3K4me3 and H3K36me3 represents a boundary between two distinct, transcriptionally-relevant chromatin states across the bodies of actively transcribed genes. These states are established (or maintained), at least in part, due to the fact that the histone methyltransferase responsible for establishing H3K4me3 binds preferentially to initiating RNA pol II associated with transcriptional start sites, while that responsible for H3K36me3 deposition binds the form of RNA pol II associated with transcriptional elongation (Bannister and Kouzarides 2011).

We found that RNA pol II exhibited significantly lower levels of enrichment at HMRs relative to up- and down-stream regions, independent of the genomic context or length of the HMR (exon/intron, 5'-/3'-proximal localization; Figures 4.3b-c, Figure C.6), and was the only ChIP feature examined to exhibit considerable negative log-fold enrichment (indicative of depletion) at HMRs. This finding is particularly striking given that RNA pol II exhibits a signal of enrichment both directly up- and downstream of HMRs. It is possible this RNA pol II depletion at HMRs is related to an alteration of RNA pol II kinetics within or surrounding highly-methylated DNA, a phenomenon observed in previous studies (Lorincz, Dickerson et al. 2004; Zilberman, Gehring et al. 2007; Maunakea, Chepelev et al. 2013). Because of the strong tendency for H3K4me3 to be highly enriched upstream of HMRs, and H3K36me3 to be highly enriched downstream of HMRs, it is tempting to speculate that, through the alteration of RNA pol II dynamics, intragenic DNA methylation plays a role in the formation of a chromatin boundary that differentiates states of transcriptional initiation and elongation within actively expressed genes. Indeed, prior studies suggest that the conversion of TSS-proximal initiating RNA pol II into the elongating form plays an important role in the establishment of the distinct chromatin state associated with gene bodies (Brookes and Pombo 2009; Badeaux and Shi 2013). Thus, our finding that RNA pol II enrichment was lowest at HMRs relative to up- and down-stream regions (Figures 4.3, Figure C.6) suggests the possibility that the strong associations seen here between DNA methylation and hPTM enrichment may result from DNA methylation's alteration of RNA pol II kinetics within and surrounding methylated DNA (Lorincz, Dickerson et al. 2004; Zilberman, Gehring et al. 2007; Maunakea, Chepelev et al. 2013).

Of all chromatin marks we investigated, only H3K4me1 consistently showed its highest levels of enrichment within HMRs relative to up- and downstream regions (where it was consistently depleted; Figures 4.3c, Figure C.5). Interestingly, while positively correlated with HMR methylation level within the HMR, we found that H3K4me1 enrichment within 1kb upstream of HMRs was *negatively* correlated with the level of HMR methylation (rho: -0.39 vs. 0.37 for 1kb upstream and within HMRs respectively; Figure 4.4). Thus, as the DNA methylation level of HMRs increases, the enrichment of H3K4me1 within those regions also increases; however, within the region directly upstream of HMRs, H3K4me1 is more depleted with increasing DNA methylation (Figures 4.4, Figure C.7). At least one recent report has noted that, within active gene bodies, H3K4me1 is important to limiting domains of H3K4me3-marked open chromatin to promoter-proximal regions (Cheng, Blum et al. 2014). Indeed, H3K4me1 is often seen flanking TSS-proximal enriched regions of H3K4me3 within active gene bodies (Kharchenko, Alekseyenko et al. 2011).

It is possible that the patterning of hPTMs around HMRs is linked to H3K4me3 exclusion, either through DNA methylation informing or being targeted to this boundary. However, we are unable to determine whether DNA methylation plays a causal role in chromatin boundary formation in insects with the current data. Nevertheless, the fact that abrupt differences in RNA pol II, H3K4 methylation, and H3K36me3 exist within and around HMRs suggests that the hypothesis that DNA methylation may alter or maintain local chromatin states warrants testing in future investigations. Notably, both the patterning of hPTMs around active gene TSSs and the alternative splicing of exons

Table 4.2. Association tests between a genomic region's differential methylation status and differential ChIP enrichment, as assessed between castes. The numbers of genomic regions falling into each pairwise category for the different hPTMs are provided along with fold enrichment of DMRs coinciding with DBRs relative to regions not differentially associated by either epigenetic signal (negative fold enrichment represents hPTM for which DMRs are under-represented among DBRs). P-values derived from a fisher's exact test.

		nonDBR	DBR	Fold enrichment	P value
H3K27ac	nonDMR	5559	1754	**1.10**	**0.0002**
	DMR	2754	980		
H3K27me3	nonDMR	10	82	**-1.23**	**0.0184**
	DMR	11	29		
H3K36me3	nonDMR	1878	2782	-1.02	NS
	DMR	1148	1607		
H3K4me1	nonDMR	1158	480	**1.24**	**0.0006**
	DMR	466	267		
H3K4me3	nonDMR	4640	3502	**1.39**	**<0.0001**
	DMR	1950	2912		
H3K9ac	nonDMR	4460	857	1.00	NS
	DMR	3349	641		
H3K9me3	nonDMR	166	104	**1.20**	**0.0386**
	DMR	172	147		
RNA Pol II	nonDMR	1266	426	-1.02	NS
	DMR	647	213		

DMR, differentially methylated region; nonDMR, non-differentially methylated region; DBR, differentially bound region (by hPTM); nonDBR, non-differentially bound region involve differences in H3K4me1 and RNA pol II (Luco, Pan et al. 2010; Luco, Allo et al. 2011; Cheng, Blum et al. 2014; Stasevich, Hayashi-Takanaka et al. 2014), thus highlighting the possibility that the regulation of genic chromatin domains may help to explain DNA methylation’s link with alternative splicing in insects (Lyko, Foret et al. 2010; Bonasio, Li et al. 2012; Herb, Wolschin et al. 2012).

Differential DNA methylation is associated with differential histone modification enrichment

We next sought to examine whether regions exhibiting significant differences in levels of DNA methylation between *C. floridanus* castes also exhibited significant differences in hPTM enrichment. Thus, we compared differentially methylated regions (DMRs) to a set of regions exhibiting significantly different hPTM enrichment (differentially bound regions: DBRs) between males and female workers.

We found that DMRs were significantly enriched for several DBRs (hPTMs H3K27ac, H3K4me1, H3K4me3, and H3K9me3) relative to methylated regions not displaying significant differences between males and workers (Table 4.2). Thus, even at the coarse resolution provided by whole body samples, DMRs exhibit significantly more DBRs than non-DMR genes.

Moreover, we found that DNA methylation biased to either males or workers was significantly associated with hPTM enrichment in the opposite phenotype for H3K4me3, and RNA pol II (Figure 4.5, Table C.5). This is again consistent with a hypothesized functional link between DNA methylation and the patterning of genic chromatin, wherein DNA methylation exhibits spatial antagonism with RNA pol II and H3K4me3.

In *Arabidopsis thaliana* (Zilberman, Coleman-Derr et al. 2008; Coleman-Derr and Zilberman 2012), and likely vertebrates (Zemach, McDaniel et al. 2010), DNA methylation is known to play a role in altering chromatin within and directly surrounding methylated regions. Specifically, methylation acts as a boundary to H2A.Z, an important TSS-associated histone variant that is linked to chromatin activation

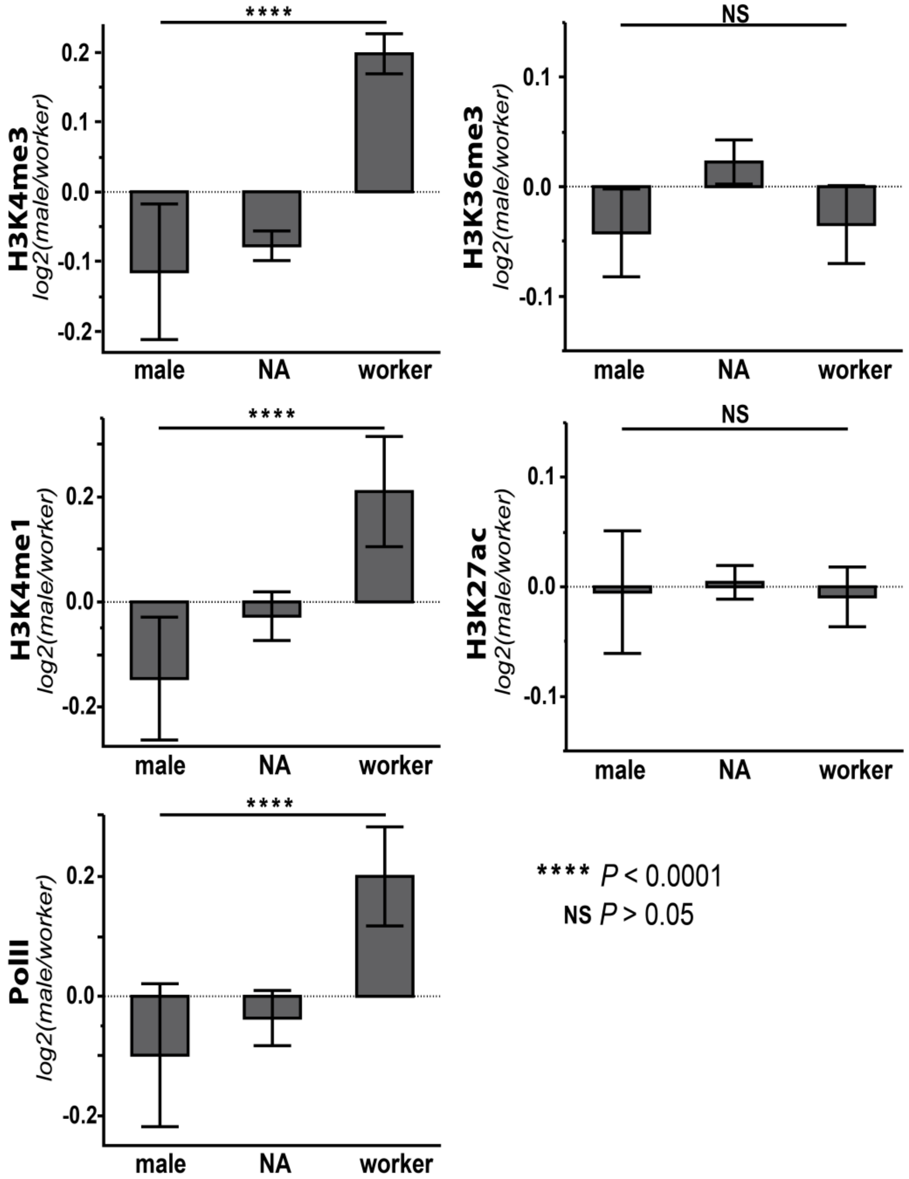


Figure 4.5. Differentially methylated regions (DMRs) show significantly different, directional hPTM enrichment between male and worker phenotypes. Active histone post-translational modification (hPTM; and RNA pol II) $\log_2$ fold differences between males andworkers as they relate to regions of significant, directional differential methylation (positive values on y axes indicate male biased ChIP enrichment, while negative values indicate worker bias; x axes: ”male”, male hypermethylated; ”worker”, worker hypermethylated; NA, not differentially methylated between phenotypes).

(Zilberman, Coleman-Derr et al. 2008; Zemach, McDaniel et al. 2010; Coleman-Derr and Zilberman 2012). Because H2A.Z is a highly-conserved component of the epigenome of active genes, and has been shown to strongly correlate with DNA methylation and promoter-proximal active gene hPTMs (Zilberman, Coleman-Derr et al. 2008), it is possible many of our observations are reflective of the conserved mechanism of H2A.Z exclusion by DNA methylation operating in insects. However, because this histone variant was not tested directly in our study, additional research will be required to test this hypothesis.

Conclusions

Our results provide several important insights into insect DNA methylation. By assessing, for the first time, the relationship between DNA methylation and hPTMs within a single insect taxon, we provide a foundation for understanding the greater epigenome in insects. In particular, our results suggest that the function of intragenic DNA methylation is linked to the function of key, active histone modifications, with over 90% of methylated genes also featuring the hPTMs H3K4me3 or H3K36me3. As additional support to this claim, we provide evidence that DNA methylation and active hPTM enrichment covary between distinct phenotypes in *C. floridanus*, suggesting that changes to DNA methylation are coupled with changes in chromatin modifications. Despite the striking concordance between DNA methylation and hPTMs, however, our results suggest the function of DNA methylation is not entirely redundant to hPTMs – DNA methylation retains explanatory power for gene expression levels when controlling for numerous hPTMs.

Studies in plants and animals have shown that variation in gene body DNA methylation affects gene regulation by altering local chromatin and the rate of elongation of RNA pol II (Zilberman, Gehring et al. 2007; Maunakea, Chepelev et al. 2013). Likewise, our findings are consistent with a functional link between DNA methylation and the organization of chromatin. Our spatial analysis of DNA methylation and hPTMs reveal a strong patterning of multiple, functionally-distinct hPTMs and RNA pol II relative to methylated regions. Most notably, RNA pol II is depleted, and H3K4me1 enriched, within highly methylated regions. We hypothesize that intragenic DNA methylation contributes to changes in chromatin and chromatin boundaries within active insect genes, particularly those that differentiate states of transcriptional initiation and elongation, occurring near the transcription start site. This hypothesis may help to explain why DNA methylation is preferentially targeted to 5'-regions of genes in most investigated insects (Bonasio, Li et al. 2012; Hunt, Glastad et al. 2013). Furthermore, as both alternative splicing and TSS-proximal chromatin organization have been linked to the dynamics of RNA pol II and H3K4me1 (among other hPTMs) (Luco, Pan et al. 2010; Luco, Allo et al. 2011; Cheng, Blum et al. 2014), it is possible that the previously observed link between DNA methylation and alternative splicing in insects (Lyko, Foret et al. 2010; Bonasio, Li et al. 2012; Herb, Wolschin et al. 2012) is influenced by hPTMs.

As we look to the future, it is clear that studies seeking to establish the epigenetic basis for developmental regulation in insects, as with environmental caste determination (Kucharski, Maleszka et al. 2008), will benefit from investigating both DNA methylation and hPTMs. In doing so, a meaningful exploration of the causal links between epigenetic modifications, chromatin boundary formation, gene regulation, and developmental fate will require extensive advancement of reverse genetic approaches to the perturbation of enzymatic mediators of epigenetic modifications in previously non-model insects.

Acknowledgements

This work was supported by the U. S. National Science Foundation (DEB-0640690) and the Georgia Tech-Elizabeth Smithgall Watts endowment.

CHAPTER 5

EFFECTS OF DNA METHYLATION AND CHROMATIN STATE ON RATES OF MOLECULAR EVOLUTION[4]

Abstract

Epigenetic information is an important regulator of gene function in eukaryotic organisms. However, epigenetic information can also influence genome evolution. Here, we investigate the importance of epigenetic marks to rates of gene evolution. We study the effects of epigenetic information on rates of molecular evolution in two disparate insects – the fly *Drosophila melanogaster* and the ant *Camponotus floridanus*, which exhibit substantial variation in DNA methylation. We found that DNA methylation was positively correlated with the synonymous substitution rate in *C. floridanus*, suggesting a key effect of DNA methylation on patterns of gene evolution. However, our data suggest that the link between DNA methylation and elevated rates of synonymous substitution was, in large part, explained by the targeting of DNA methylation to genes with signatures of transcriptionally-active chromatin, rather than the mutational effect of DNA methylation itself. This result suggests that chromatin structure, rather than the mutational effects of DNA methylation, may be the primary epigenetic driver of genome evolution in insects. This phenomenon may be explained by an elevated mutation rate for genes residing in transcriptionally active chromatin, or by increased structural constraint on genes in inactive chromatin. Overall, our study highlights how different epigenetic systems contribute to variation in the rates of coding sequence evolution.

Introduction

[4] Glastad, K. M., M. A. D. Goodisman, et al. 2015. Effects of DNA methylation and chromatin state on rates of molecular evolution in insects. G3.

The evolutionary rates of protein-coding genes span multiple orders of magnitude within the genome of a single taxon (Wolf, Novichkov et al. 2009). Determining the functional, structural, and regulatory sources of variation in constraints on protein-coding sequences has been central to advancing our understanding of evolution at the molecular level (Koonin and Wolf 2010). Accordingly, a large and growing body of research has revealed fundamental insights into near-universal constraints on protein-coding sequence evolution (Pal, Papp et al. 2006; Koonin and Wolf 2010). These constraints include the essentiality of a protein to organismal survival (Wall, Hirsh et al. 2005; Liao, Scott et al. 2006), gene expression level (Drummond and Wilke 2008), gene expression pattern (Duret and Mouchiroud 2000; Hunt, Ometto et al. 2013), and gene compactness (Eisenberg and Levanon 2003; Carmel, Rogozin et al. 2007).

In addition, chromatin structure has recently been investigated as a factor influencing molecular evolution. Associations between chromatin structure and constraints on gene evolution can arise as a byproduct of the link between chromatin structure and gene expression patterns (Prendergast, Campbell et al. 2007; Filion, van Bemmel et al. 2010; Kharchenko, Alekseyenko et al. 2011). Variation in mutation rate and sequence constraints are also linked to nucleosome positioning and chromatin accessibility (Prendergast, Campbell et al. 2007; Prendergast and Semple 2011; Tolstorukov, Volfovsky et al. 2011; Schuster-Bockler and Lehner 2012; Langley, Karpen et al. 2014; Makova and Hardison 2015). However, this issue has yet to be investigated in insect genomes, where evolutionary variation in DNA methylation (Glastad, Hunt et al. 2011) provides the opportunity to disentangle the relative effects of DNA methylation and other epigenetic marks.

Our primary interest in undertaking this study was to better understand how DNA methylation and chromatin structure affect genome evolution. Studies in plants and animals have shown that variation in intragenic DNA methylation affects gene regulation by altering local chromatin and the rate of elongation of RNA pol II (Zilberman, Gehring et al. 2007; Maunakea, Chepelev et al. 2013). Similarly, the regulatory roles of histone modifications are known to include the mediation of binding affinities of protein complexes, such as those related to transcriptional and splicing machinery, as well as the direct alteration of local chromatin structure (Bintu, Ishibashi et al. ; Luco, Pan et al. 2010; Bell, Tiwari et al. 2011). Together, DNA methylation and histone modifications interact to contribute to a multi-faceted epigenetic landscape in eukaryotic cells (Cedar and Bergman 2009). For example, in insects with functional DNA methylation systems, the targeting of DNA methylation has been shown to exhibit striking associations with multiple histone modifications that are, in turn, linked to active transcription (Nanty, Carbajosa et al. 2011; Hunt, Glastad et al. 2013; Glastad, Hunt et al. 2015).

Although DNA methylation is predominantly targeted to cytosines at cytosine-phosphate-guanine (CpG) dinucleotides in eukaryotes (Klose and Bird 2006), the localization of DNA methylation varies substantially among taxa. In vertebrate animals, DNA methylation is present globally within the genome, with only small regions of unmethylated DNA found largely in gene promoters (Suzuki and Bird 2008). In contrast, the genomes of invertebrates exhibit relatively sparse levels of DNA methylation, present almost exclusively in genes(Suzuki and Bird 2008; Feng, Cokus et al. 2010; Zemach, McDaniel et al. 2010) (Suzuki and Bird 2008; Feng, Cokus et al. 2010; Zemach, McDaniel et al. 2010). DNA methylation is known to increase the mutation rate of affected cytosines (Bird 1980; Elango, Kim et al. 2008; Mugal and Ellegren 2011; Drewell, Bush et al. 2014). Despite this mutational effect, however, the presence of DNA methylation in gene bodies is paradoxically associated with protein conservation (Takuno and Gaut 2012; Chuang and Chiang 2014; Glastad, Hunt et al. 2014). Thus, the effects and associations of DNA methylation, with respect to gene evolution, remain nebulous.

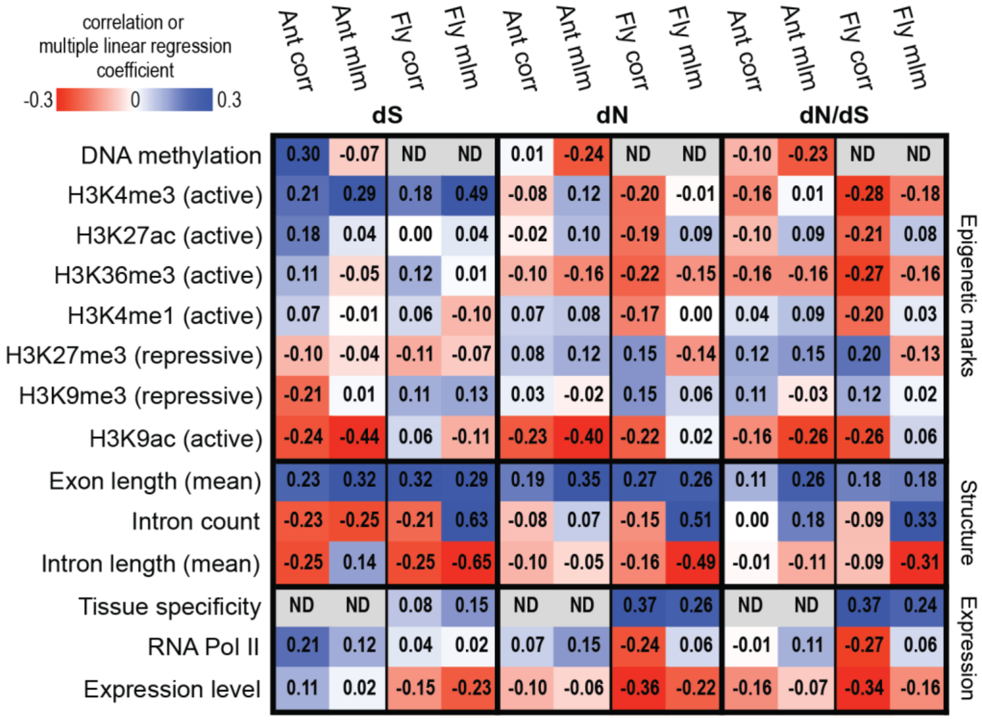
Here, we investigate the relationships between epigenetic marks and coding sequence evolution in two insects, the fruit fly *Drosophila melanogaster* and the carpenter ant *Camponotus floridanus*. Distinct chromatin states have been well characterized in *D. melanogaster* (Filion, van Bemmel et al. 2010; Kharchenko, Alekseyenko et al. 2011) and, more recently, genome-wide spatial profiles of many histone modifications have been examined in the ant *C. floridanus* (Simola, Ye et al. 2013) . Importantly, a comparison of these taxa provides a novel opportunity to determine the contribution of DNA methylation to coding sequence evolution because *C. floridanus* exhibits substantial genomic DNA methylation (Bonasio, Li et al. 2012) but *D. melanogaster* does not (Zemach, McDaniel et al. 2010; Takayama, Dhahbi et al. 2014). Therefore, our investigation allows us to isolate the effects of DNA methylation on gene evolution and provide direct insight into how epigenetic information affects molecular evolution in eukaryotes.

Results and Discussion

Coding sequence evolution in the presence and absence of DNA methylation:

Our first goal in this study was to understand how DNA methylation affects rates of molecular evolution. In *C. floridanus*, we observed that DNA methylation was the second largest negative correlate of both dN and dN/dS, when controlling for other correlates of substitution rate using multiple linear regression models (Figs. 5.1 and D.2). This association is consistent with the preferential targeting of DNA methylation to constitutively expressed, phylogenetically conserved genes in insect genomes (Sarda, Zeng et al. 2012; Hunt, Glastad et al. 2013; Glastad, Hunt et al. 2014).

In contrast, we observed a strong positive correlation between DNA methylation and dS in *C. floridanus* (Figs. 5.1 and D.1). In line with this finding, there is ample evidence that DNA methylation results in elevated mutation rates in mammals (Elango, Kim et al. 2008; Mugal and Ellegren 2011), as well as in many insects (Glastad, Hunt et al. 2011; Glastad, Hunt et al. 2013; Drewell, Bush et al. 2014), which exhibit much lower levels of DNA methylation than vertebrates (Zemach, McDaniel et al. 2010). Thus, the



	dS				dN				dN/dS				
	Ant corr	Ant mlm	Fly corr	Fly mlm	Ant corr	Ant mlm	Fly corr	Fly mlm	Ant corr	Ant mlm	Fly corr	Fly mlm	
DNA methylation	0.30	-0.07	ND	ND	0.01	-0.24	ND	ND	-0.10	-0.23	ND	ND	Epigenetic marks
H3K4me3 (active)	0.21	0.29	0.18	0.49	-0.08	0.12	-0.20	-0.01	-0.16	0.01	-0.28	-0.18	
H3K27ac (active)	0.18	0.04	0.00	0.04	-0.02	0.10	-0.19	0.09	-0.10	0.09	-0.21	0.08	
H3K36me3 (active)	0.11	-0.05	0.12	0.01	-0.10	-0.16	-0.22	-0.15	-0.16	-0.16	-0.27	-0.16	
H3K4me1 (active)	0.07	-0.01	0.06	-0.10	0.07	0.08	-0.17	0.00	0.04	0.09	-0.20	0.03	
H3K27me3 (repressive)	-0.10	-0.04	-0.11	-0.07	0.08	0.12	0.15	-0.14	0.12	0.15	0.20	-0.13	
H3K9me3 (repressive)	-0.21	0.01	0.11	0.13	0.03	-0.02	0.15	0.06	0.11	-0.03	0.12	0.02	
H3K9ac (active)	-0.24	-0.44	0.06	-0.11	-0.23	-0.40	-0.22	0.02	-0.16	-0.26	-0.26	0.06	
Exon length (mean)	0.23	0.32	0.32	0.29	0.19	0.35	0.27	0.26	0.11	0.26	0.18	0.18	Structure
Intron count	-0.23	-0.25	-0.21	0.63	-0.08	0.07	-0.15	0.51	0.00	0.18	-0.09	0.33	
Intron length (mean)	-0.25	0.14	-0.25	-0.65	-0.10	-0.05	-0.16	-0.49	-0.01	-0.11	-0.09	-0.31	
Tissue specificity	ND	ND	0.08	0.15	ND	ND	0.37	0.26	ND	ND	0.37	0.24	Expression
RNA Pol II	0.21	0.12	0.04	0.02	0.07	0.15	-0.24	0.06	-0.01	0.11	-0.27	0.06	
Expression level	0.11	0.02	-0.15	-0.23	-0.10	-0.06	-0.36	-0.22	-0.16	-0.07	-0.34	-0.16	

Figure 5.1. Correlation coefficients (corr) and multiple linear regression model coefficients (mlm) between sequence substitution rates and gene characteristics in the ant *Camponotus floridanus* and the fly *Drosophila melanogaster*. 'Active' and 'repressive' histone modification designations indicate associations with active transcription and repression of transcription in *D. melanogaster* (Kharchenko et al. 2011). *C. floridanus* n = 4984 genes, *D. melanogaster* n = 7396 genes. Abbreviations: H3, histone H3; K, lysine; ac, acetylation; me1, monomethylation; me3, trimethylation; Pol, Polymerase.

most parsimonious explanation for the positive correlation between DNA methylation and dS would appear to be that the increased mutability of methylated cytosines has resulted in an elevated rate of synonymous substitutions at methylated loci.

Surprisingly, the positive correlation between DNA methylation and dS did not persist when controlling for other factors in our multiple linear regression analysis (Figure 5.1). This suggested that further investigation was needed to assess whether DNA methylation is in fact the primary causal factor underlying its positive correlation with synonymous substitution rate in *C. floridanus*. Therefore, we sought to determine whether the correlation between dS and DNA methylation was the consequence of substitutions at CpG dinucleotides, where DNA methylation is predominantly targeted in animal genomes (Bird 1980; Bonasio, Li et al. 2012).

We assessed correlations between DNA methylation in *C. floridanus* and dS among ants after masking positions with a CpG dinucleotide in any of the taxa included in multiple sequence alignments. Based on the hypothesis that DNA methylation causes an increase in both mutation rate and the rate of synonymous substitution, we predicted that we would not detect a significant correlation between DNA methylation and dS after masking CpG dinucleotides. Although the masking of CpG sites did indeed reduce the strength of correlation between dS and DNA methylation by 52%, a positive correlation between DNA methylation and dS persisted (Figure 5.2). One possible explanation for this finding is that neighboring methylated sites are subject to elevated mutation rates (Qu, Hashimoto et al. 2012). However, the masking of CpG sites resulted in a reduction in correlations between dS and every factor we investigated in this study, by an average of 47% (Table D.1). Thus, we sought to gain further insight into the cause of the positive correlation between intragenic DNA methylation and dS by testing for an association between DNA methylation in *C. floridanus* and dS measured in *Drosophila* orthologs.

We predicted there would be no significant association between DNA methylation, as measured in *C. floridanus*, and orthologous dS measured among only

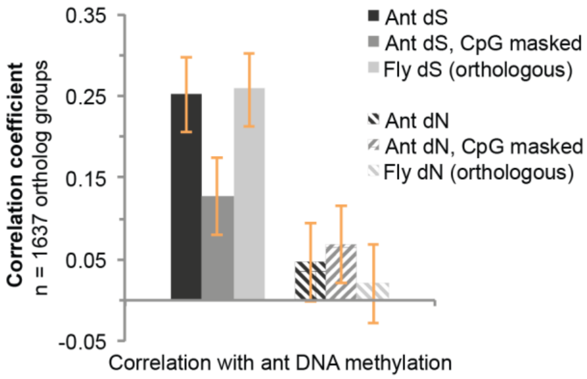


Figure 5.2. Correlations between *C. floridanus* DNA methylation and sequence substitution rates of ortholog groups in either ants or flies. Pearson's correlation coefficients with 95% confidence intervals are shown.

Drosophila species, because DNA methylation does not exist at substantial levels in the genome of *D. melanogaster* or other flies (Urieli-Shoval, Gruenbaum et al. 1982; Zemach, McDaniel et al. 2010). Surprisingly, however, the strength of the positive correlation between *C. floridanus* DNA methylation and dS among ants did not differ significantly from the strength of the correlation between *C. floridanus* DNA methylation and dS calculated solely among *Drosophila* orthologs (Figure 5.2). This result provided evidence that DNA methylation is unlikely to be the dominant causal factor driving the elevated rate of synonymous substitutions observed for methylated genes in our study.

The possibility that other processes, besides DNA methylation, were responsible for the observed correlations with dS in insects is bolstered by an analysis of DNA methylation and substitution rate in introns of *Homo sapiens*, which revealed that DNA methylation level co-varies with other factors that influence the overall substitution rate (Mugal and Ellegren 2011). However, in *H. sapiens*, DNA methylation was found to exhibit a strong influence on the CpG transition rate (Mugal and Ellegren 2011). We note that a more limited role in shaping variation in mutation rates may be expected for DNA methylation in insects and other invertebrates, as compared to vertebrate taxa, for at least two reasons. First, invertebrates exhibit substantially lower levels of DNA methylation than vertebrates (Zemach, McDaniel et al. 2010). Second, DNA methylation is often selectively localized to the 5'-region of genes in holometabolous insect taxa (Bonasio, Li et al. 2012; Hunt, Glastad et al. 2013), while DNA methylation is globally targeted in the genomes of vertebrates (Suzuki and Bird 2008). What, then, is responsible for the elevated rates of synonymous substitutions observed for methylated genes in insects?

DNA methylation is linked to chromatin states affecting coding sequence evolution

Recent studies have revealed that DNA methylation is integrated into domains of transcriptionally active chromatin in insect genomes (Nanty, Carbajosa et al. 2011; Hunt, Glastad et al. 2013; Glastad, Hunt et al. 2015). Thus, we chose to investigate whether combinatorial epigenetic states may explain the observed associations between coding sequence evolution and DNA methylation in *C. floridanus*. To this end, we performed a principal component analysis (PCA) of DNA methylation and seven histone modifications in *C. floridanus*, as well as another PCA of the same seven histone modifications in *D. melanogaster*. These analyses provided proxies for the assessment of distinct chromatin states among coding sequences.

Three principal components (PCs) in each taxa explained greater than 10% of total variance in epigenetic marks, and the top three PCs together explained 76% and 82% of the total epigenetic variance in *C. floridanus* and *D. melanogaster*, respectively (Table 5.1). Among the top three PCs, DNA methylation loaded most heavily on PC1 in *C. floridanus* (Table 5.1). *C. floridanus* PC1 explained 39% of the total variance in epigenetic marks and exhibited relatively large positive loadings of DNA methylation and three histone modifications associated with active transcription: H3K4me3, H3K36me3, and H3K27ac ("active" modifications; (Kharchenko, Alekseyenko et al. 2011)). Similarly, *D. melanogaster* PC1 explained 55% of the total variance in epigenetic marks and also exhibited large positive loadings of these active histone

Table 5.1. Principal component (PC) analysis of epigenetic marks illustrate associations between chromatin state and coding sequence evolution

	***C. floridanus* (ant)**			***D. melanogaster* (fly)**		
	PC1 (39.4%)	PC2 (20.0%)	PC3 (16.6%)	PC1 (55.2%)	PC2 (14.8%)	PC3 (11.7%)
	Eigenvectors					
DNA methylation	0.45	-0.33	-0.04	ND	ND	ND
H3K4me3	0.46	0.36	0.03	0.46	0.21	-0.12
H3K27ac	0.45	0.22	0.00	0.37	0.41	-0.16
H3K36me3	0.42	0.04	0.14	0.40	-0.33	0.40
H3K4me1	0.21	-0.39	0.45	0.37	-0.22	0.36
H3K27me3	0.02	0.18	0.77	-0.40	0.14	-0.25
H3K9ac	-0.04	0.72	-0.03	0.41	0.34	-0.41
H3K9me3	-0.40	0.07	0.43	-0.17	0.71	0.67
	Correlation coefficients of gene expression metrics with PCs					
RNA Pol II	0.60****	0.07**	0.19****	0.78****	0.25****	-0.08***
Expression level	0.59****	0.01	-0.04	0.55****	0.02	0.00
Tissue specificity	ND	ND	ND	-0.60****	-0.05*	-0.01
	Correlation coefficients of sequence substitution metrics with PCs					
dS	0.21****	-0.11****	-0.11****	0.20****	0.05*	0.10****
dN	-0.03	-0.15****	0.06**	-0.05*	-0.05*	0.07***
dN/dS	-0.11****	-0.12****	0.11****	-0.13****	-0.07***	0.04

*P < 0.05, **P < 10-2, ***P < 10-3, ****P < 10-4; ND, no data.

PCs explaining less than 10% variation are not shown.

modifications. In contrast, H3K9ac, which differed in its association with transcription in *C. floridanus* and *D. melanogaster* (Figure D.3), negatively loaded on *C. floridanus* PC1 and positively loaded on *D. melanogaster* PC1. The histone modifications H3K9me3 and H3K27me3, which are associated with low transcriptional activity ("repressive" modifications; (Kharchenko, Alekseyenko et al. 2011)), both loaded negatively onto *D. melanogaster* PC1, while only H3K9me3 loaded negatively on *C. floridanus* PC1.

We found that PC1 exhibited striking positive correlations with both gene expression level and RNA Pol II occupancy in both taxa (Table 5.1). We also found that genes with high values of *C. floridanus* PC1 were significantly enriched for gene ontology biological process terms related to cellular housekeeping functions, including 'ribosome biogenesis', 'translation', and 'proton transport' (Table D.2). Accordingly, large PC1 values can be thought of as representing a transcriptionally active chromatin state in both taxa.

PC1 was also positively correlated with dS in both *C. floridanus* and *D. melanogaster* (Table 5.1). The positive correlation between PC1 and dS, coupled with the integration of DNA methylation into *C. floridanus* PC1, suggests that a transcriptionally active or "open" chromatin state may explain the bulk of the observed positive correlation between DNA methylation and dS in *C. floridanus* (Table 5.1).
To further investigate the hypothesis that chromatin state was the critical factor affecting variation in rates of evolution in synonymous sites, we again leveraged the evolutionary loss of DNA methylation in *D. melanogaster*. We predicted that histone modifications in the genome of *D. melanogaster* that (i) are markers of transcriptionally active chromatin and (ii) are highly correlated with DNA methylation in the genome of *C. floridanus*, would be positively correlated with dS measures among *Drosophila* species. Thus, we tested whether histone modifications that are correlated with DNA methylation levels in *C. floridanus* (Fig 5.3a; (Glastad, Hunt et al. 2015)) were also correlated with dS in *D. melanogaster,* despite its absence of DNA methylation.

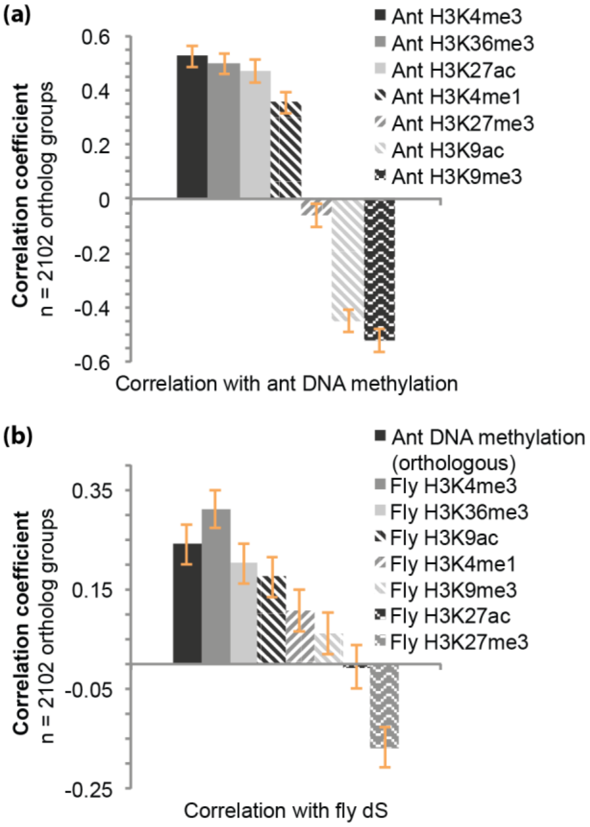


Figure 5.3. Correlations between DNA methylation and synonymous sequence substitution are mirrored by several histone modifications in insect genomes. (a) Correlations between histone modifications and DNA methylation in the ant *C. floridanus*. (b) Correlations between histone modifications in the fly *D. melanogaster* and sequence substitution in flies mirror the relationship between *C. floridanus* DNA methylation and orthologous sequence substitution in flies. Pearson's correlation coefficients with 95% confidence intervals are shown.

Remarkably, the two histone modifications that were most strongly correlated with DNA methylation in *C. floridanus*, H3K4me3 and H3K36me3, exhibited correlations with dS in *Drosophila* orthologs that did not differ significantly from the correlation between *Drosophila* dS and DNA methylation in *C. floridanus* orthologs (Figure 5.3b). We interpret this result as support for the hypothesis that loci residing in conserved, transcriptionally active chromatin domains (Engström, Ho Sui et al. 2007; Hunt, Glastad et al. 2013) exhibit elevated rates of synonymous substitution in insect genomes, irrespective of the presence or absence of DNA methylation.

These findings raise the question of why genes residing in transcriptionally active chromatin would exhibit elevated synonymous substitution rates. One possible explanation is that genes residing in transcriptionally active chromatin exhibit elevated mutation rates resulting from the process of transcription itself. In support of this idea, a study of single-celled yeast and human germline cells recently revealed that mutation rates are positively correlated with gene expression level (Park, Qian et al. 2012). This suggests that eukaryotic transcription exerts a net mutagenic effect, in spite of transcription-coupled repair. Another possible explanation for elevated rates of synonymous substitution in regions of active chromatin is that selection acts more strongly on synonymous sites in regions of inaccessible chromatin than accessible chromatin, as suggested by an analysis of chromatin states and molecular evolution in *H. sapiens* (Prendergast, Campbell et al. 2007).

Conclusions

We investigated how epigenetic marks, transcription, and gene structure relate to substitution rates in the genes of two highly diverged insect taxa. We found that DNA methylation was positively correlated with the rate of synonymous substitution. However, by comparing processes of molecular evolution in the presence and absence of DNA methylation, we revealed that this relationship was not explained primarily by the mutability of methylated cytosines in insects. Instead, the relationship between DNA methylation and synonymous substitution was apparently explained in large part by the targeting of DNA methylation to genes with signatures of transcriptionally active chromatin. We hypothesize that active chromatin may be prone to elevated rates of synonymous substitution by way of mutational pressures imposed by active transcription, or by differences in the structural requirements of distinct chromatin states. Overall, this research provides new insights into how epigenetic factors affect genome evolution in insects and other eukaryotic systems.

Material and Methods

Molecular evolution

Single-copy orthology was assigned (i) across seven ant species (*C. floridanus, Harpegnathos saltator*, *Linepithema humile*, *Pogonomyrmex barbatus*, *Solenopsis invicta*, *Acromyrmex echinator*, and *Atta cephalotes*) and (ii) between *C. floridanus* and *D. melanogaster* by orthoDB (Waterhouse, Zdobnov et al. 2011; Simola, Wissler et al. 2013).

Multiple sequence alignment was performed with PRANK (Löytynoja and Goldman 2005), as implemented by GUIDANCE (Penn, Privman et al. 2010). PhyML (Guindon, Dufayard et al. 2010) was used to impute trees from multiple sequence alignments, modifying branch lengths and rate variables, but keeping topology the same as input trees. Gblocks (Talavera and Castresana 2007) was then used to filter alignment columns, using default settings, prior to further analyses.

Coding sequence substitution rates for *D. melanogaster*, as summed over species from the *Drosophila melanogaster* species subgroup, were calculated previously (Clark, Eisen et al. 2007). Substitution rates for ants were averaged across all aligned codons for a given protein, with free dN/dS ratios for each branch, using PAML with the F3x4 codon model (Yang 2007). We filtered out genes for which dN or dS values were greater than 14 across the 7 ant tree, as well as genes that had an aligned length of less than 50 codons. In order to mask CpG dinucleotides for an additional analysis, a separate dataset was produced wherein alignment columns with a CpG in any of the aligned species were masked before running PAML.

Chromatin immunoprecipitation sequencing (ChIP-seq)

We used ChIP-seq data that were generated previously for *C. floridanus* (Simola, Ye et al. 2013). We remapped these data to the *C. floridanus* genome (Cflo_3.3) using bowtie (Langmead, Trapnell et al. 2009) after filtering for adapter contamination and read quality using Trimmomatic (Bolger, Lohse et al. 2014). We allowed one mismatch in the "seed" region and only accepted the most valid alignment for each mapping read.

MACS2 (Zhang, Liu et al. 2008) was then used to estimate the read enrichment relative to an input control (as well as bulk histone H3 profiles for histone modifications to histone H3) for each ChIP library after removal of duplicate reads. We only allowed one of each duplicated read when running MACS in an effort to minimize bias introduced through PCR amplification. ChIP enrichment scores were assigned to a coding sequence (CDS) as fold enrichment value over normalized read counts overlapping the given CDS for merged libraries from major workers, minor workers, and males (Simola, Ye et al. 2013).

ChIP-seq data from *D. melanogaster* embryos were obtained for each histone modification from modEncode ((Celniker, Dillon et al. 2009); modENCODE ID numbers: 3955, 4120, 4938, 4939, 4950, 5092, 5096, 5103), and mapped to *D. melanogaster* genome build r5.42 CDS annotations. *D. melanogaster* ChIP enrichment scores were assigned to a coding sequence (CDS) following the methods described for *C. floridanus*.

Whole genome bisulfite sequencing (WGBS)

We calculated fractional DNA methylation levels, as averaged across all CpG dinucleotides from a given coding sequence, following methods described in detail previously (Hunt, Glastad et al. 2013). We used previously-generated WGBS data from *C. floridanus* (Bonasio, Li et al. 2012), accessed from the NCBI GEO database (GSE31577). DNA methylation levels were assessed for merged libraries from queens, workers, and males.

Transcriptome sequencing (RNA-seq)

RNA-seq reads from adult *C. floridanus* were generated previously (Bonasio, Li et al. 2012). We filtered (Bolger, Lohse et al. 2014) and aligned these reads to the *C. floridanus* genome (v3.3) using tophat (Trapnell, Pachter et al. 2009). Cufflinks (Roberts, Trapnell et al. 2011) was then run with multi-read-correction, fragment bias correction, and upper quartile normalization. Cuffdiff (Roberts, Trapnell et al. 2011) fpkm values from queen, worker, and male libraries were averaged to represent *C. floridanus* gene expression level.

We used *D. melanogaster* RNA-seq 'modENCODE Transcriptome v2 Expression Scores', obtained from the Berkeley Drosophila Genome Project (http://fruitfly.org/sequnce/download.html; (Celniker, Dillon et al. 2009)). The mean of gene expression levels from four day post-eclosion mated male and female heads was used to represent *D. melanogaster* gene expression level.

Gene structure and annotation

Mean intron and exon sizes were calculated using *C. floridanus* 3.3 gene models and *D. melanogaster* flybase v5.42 (FB2011_10) gene models.

C. floridanus gene ontology (GO) annotations were assigned using Blast2GO (Conesa, Gotz et al. 2005). Blast2GO's inbuilt 'gossip' package was used to test for enrichment using a Fisher's exact test, correcting for multiple testing using a Benjamini-Hochberg false discovery rate (FDR). Significantly enriched terms (FDR $P < 0.05$) were reduced to the most specific enriched terms for presentation.

Statistical analyses

Prior to linear model analysis, all data were log-transformed (following the addition of 0.0001 to prevent discarding zero values) and then standardized (mean = 0, standard deviation = 1) in the R statistical computing environment (R Development Core Team 2011). Multiple linear regression models were fitted with the 'lm' function in R, and confidence intervals for model parameters were obtained with the 'confint' function in R.

The JMP statistical software package (SAS Institute Inc, Cary, NC) was used to perform principal component analysis, which directly addresses the issue of collinearity among variables, and to calculate Pearson's correlations. We found that multiple linear regression models using substitution rates summed over seven ant species explained greater variance in dependent variables than those measured for the *C. floridanus* branch alone (seven ant dS R^2 = 0.27, *C. floridanus* branch dS R^2 = 0.11; seven ant dN R^2 = 0.19, *C. floridanus* branch dN R^2 = 0.17). Thus, we chose to use dN and dS values summed over the seven ant tree.

Acknowledgements

This work was supported by the Georgia Tech–Elizabeth Smithgall Watts endowment.

CHAPTER 6

CONCLUSIONS

This dissertation encompasses four studies focused on understanding the molecular basis of caste formation in social insects. These studies focused on studying the impact of DNA methylation on the epigenetic production of castes in social insects, as well as the epigenomic context, and the evolutionary correlates of insect DNA methylation. From these analyses, we found that DNA methylation serves an important role in transcriptionally active insect genes, shows a complex, but present association with alternative phenotype, and functionally interacts with other caste-related epigenetic signals.

The results from chapter two, *Epigenetic inheritance and genome regulation: is DNA methylation linked to ploidy in haplodiploid insects,* suggest that DNA methylation is associated with determining caste in the fire ant *Solenopsis invicta*, and further suggests a role for DNA methylation in compensating for differences in ploidy between haplodiploid insect sexes. In hymenopteran social insects, sex is determined by offspring ploidy, with haploid individuals developing into males, and diploid individuals developing into females. In some hymenopteran insects including the fire ant however, diploid individuals can sometimes develop into males. By including both haploid and diploid males in our study, we were able to identify that the greatest number of differences between phenotypes in our study existed between males of differing ploidy level (but highly similar phenotype), suggesting a novel role for DNA methylation in mediating molecular compensation for ploidy differences between sexes of haplodiploid species.

The results of chapter three, *The caste- and sex- specific methylome of the termite Zootermopsis nevadensis*, show that DNA methylation is strongly associated with termite caste, and targets more genes than seen in other social insects. We further find that differentially methylated loci are actually less variably expressed between castes than methylated genes that do not differ. However, differentially methylated genes do exhibit higher levels of alternative splicing, and are strongly enriched for multiple TF regulatory motifs, as well as mi-RNA profiles. These data are consistent with a primary role of intragenic DNA methylation in dampening gene expression noise at key loci (as suggested in (Huh, Zeng et al. 2013)), and suggests that differential methylation may play a similar role at genes with phenotype-specific increased susceptibility to expression noise due to other regulatory differences between phenotypes. This may explain at least in part the cryptic nature of differential methylation relative to transcription observed in social insect, despite the former's connection to caste determination.

Chapter four, *The epigenomic context of insect DNA methylation,* demonstrates that methylated regions of insect genes show distinct chromatin signatures. Furthermore, we find that differences in DNA methylation and several important histone modifications covary. These results both integrate DNA methylation into our understanding of other chromatin modifications in insects, as well as highlight a potential mechanism through which DNA methylation may mediate alternative phenotype (through an interaction with active chromatin).

The results from chapter five, *Effects of DNA methylation and chromatin state on the rates of molecular evolution in insects*, elucidate the molecular evolutionary associations of insect DNA methylation, and provide further context to its distribution and targeting in insects. By comparing DNA methylation to signals of directed and neutral evolution in insects, we were able to identify several important correlates of DNA methylation that shed light on its evolutionary context. Furthermore, by integrating evolutionary data from an insect that lacks DNA methylation as well as other epigenetic signals in both species, we were able to link DNA methylation's putative impact on evolutionary rate to more general transcriptomic factors. This research furthers the emerging understanding that epigenetic and transcriptional status greatly impacts the mutational and evolutionary capacity of genes (Makova and Hardison 2015).

This research has demonstrated that DNA methylation is associated with alternative phenotypes in multiple social insects. However, the association between phenotype and DNA methylation is complex. The association between DNA methylation and caste may be more important in the under-studied hemipteran insects such as the termite, where it is much more widely distributed among genes. DNA methylation seems to be preferentially targeted to more conserved, less variably expressed genes, with genes showing DNA methylation differences being some of the *least* variably expressed genes between castes. Furthermore, these associations are potentially underlain by close co-targeting of DNA methylation to regions associated with active transcriptional elongation. This is particularly pronounced in hymenopteran social insects, where gene-start-proximal DNA methylation is strongly localized to gene regions flanking promoters, where ChIP-sequencing data suggests RNA polymerase II is actively transitioning from its initiating to its elongating form (Hunt, Glastad et al. 2013). Finally, this research shows that previously-observed molecular evolutionary associations with DNA methylation also exist in *D. melanogaster* (which lacks DNA methylation), suggesting DNA methylation's association with molecular evolutionary rate is underwritten by more-conserved, co-associated epigenomic features in insects.

It is tempting to speculate that, given these results, a major component of insect DNA methylation's role in alternative phenotype definition is to buffer the transcriptome at genes where phenotype-specific regulatory changes (eg TF-binding/expression, chromatin changes) would otherwise lead to unacceptable expression noise due to increased spurious DNA binding. Due to the rapid drop in sequencing costs associated with advances in technology, it is likely that well-informed future studies will be able to disentangle this complex relationship, and provide further insight into the epigenetic foundations for social insect caste determination.

APPENDIX A

SUPPLEMENTARY MATERIAL FOR CHAPTER 2

Supplementary Tables and Figures

Table A.1. Summary of non-conversion statistics from an unmethylated spike-in control (enterobacteria phage lambda DNA, GenBank accession J02459.1)

Caste	CG sites	Mean fractional methylation	Mean coverage	FDR "methylated" sites
A	3110	0.004075	553.272	5
MD	3112	0.003598	578.485	5
MH	3112	0.003576	600.945	4
W	3111	0.003723	584.498	6

Table A.2. Enrichment of GO annotations among DMGs (relative to non-DMGs) - all P < 0.05

GO-ID	Term	Category	FDR	P-Value	#Test	#Ref	#notAnnotTest	#notAnnotRef
GO:0005488	binding	F	**0.0001**	6.55E-07	1434	868	658	565
GO:0000166	nucleotide binding	F	**0.0277**	0.0003	464	249	1628	1184
GO:0032502	developmental process	P	**0.0435**	0.0008	278	140	1814	1293
GO:0005694	chromosome	C	**0.0435**	0.0008	122	50	1970	1383
GO:0032501	multicellular organismal process	P	0.0686	0.0020	255	130	1837	1303
GO:0007275	multicellular organismal development	P	0.0686	0.0020	255	130	1837	1303
GO:0048856	anatomical structure development	P	0.0830	0.0032	152	71	1940	1362
GO:0009653	anatomical structure morphogenesis	P	0.0830	0.0032	152	71	1940	1362
GO:0016787	hydrolase activity	F	0.1045	0.0045	459	262	1633	1171
GO:0003676	nucleic acid binding	F	0.1450	0.0084	548	324	1544	1109
GO:0009056	catabolic process	P	0.1450	0.0092	275	150	1817	1283
GO:0030154	cell differentiation	P	0.1450	0.0093	145	71	1947	1362
GO:0048869	cellular developmental process	P	0.1450	0.0093	145	71	1947	1362
GO:0003824	catalytic activity	F	0.1450	0.0114	1042	657	1050	776
GO:0004672	protein kinase activity	F	0.1450	0.0121	95	43	1997	1390
GO:0016773	phosphotransferase activity, alcohol group as acceptor	F	0.1450	0.0121	95	43	1997	1390
GO:0016301	kinase activity	F	0.1450	0.0125	148	74	1944	1359
GO:0016772	transferase activity, transferring phosphorus-containing groups	F	0.1450	0.0125	148	74	1944	1359
GO:0005634	nucleus	C	0.1516	0.0138	536	320	1556	1113
GO:0003677	DNA binding	F	0.1908	0.0218	206	112	1886	1321
GO:0031975	envelope	C	0.1908	0.0219	31	10	2061	1423
GO:0031967	organelle envelope	C	0.1908	0.0219	31	10	2061	1423
GO:0012505	endomembrane system	C	0.1908	0.0219	31	10	2061	1423
GO:0005635	nuclear envelope	C	0.1908	0.0219	31	10	2061	1423
GO:0003700	sequence-specific DNA binding transcription factor activity	F	0.2090	0.0272	53	22	2039	1411
GO:0001071	nucleic acid binding transcription factor activity	F	0.2090	0.0272	53	22	2039	1411
GO:0006259	DNA metabolic process	P	0.2090	0.0280	117	59	1975	1374
GO:0090304	nucleic acid metabolic process	P	0.2090	0.0280	117	59	1975	1374
GO:0065007	biological regulation	P	0.2432	0.0337	645	400	1447	1033
GO:0008289	lipid binding	F	0.2552	0.0366	31	11	2061	1422
GO:0050789	regulation of biological process	P	0.2956	0.0438	635	396	1457	1037

P, biological process; F, molecular function; C, cellular component

#Test, number of genes with the designated annotation in test set; #Ref, number of genes with the designated annotation in reference set; #notAnnotTest, number of genes lacking the designated annotation in test set; #notAnnotRef, number of genes lacking the designated annotation in reference set

Table A.3. Enrichment of GO annotations among non-DMGs (relative to DMGs) - all P < 0.05

GO-ID	Term	Category	FDR	P-Value	#Test	#Ref	#notAnnotTest	#notAnnotRef
GO:0005198	structural molecule activity	F	**0.0082**	3.93E-05	84	64	1349	2028
GO:0044444	cytoplasmic part	C	**0.0103**	9.86E-05	385	447	1048	1645
GO:0030529	ribonucleoprotein complex	C	**0.0427**	8.37E-04	115	111	1318	1981
GO:0005840	ribosome	C	**0.0427**	8.37E-04	115	111	1318	1981
GO:0006412	translation	P	**0.0488**	0.0016	138	138	1295	1892
GO:0044249	cellular biosynthetic process	P	0.1074	0.0041	129	142	1304	1958
GO:0009059	macromolecule biosynthetic process	P	0.1074	0.0041	129	142	1304	1958
GO:0034645	cellular macromolecule biosynthetic process	P	0.1074	0.0041	129	142	1304	1958
GO:0005730	nucleolus	C	0.1561	0.0067	62	57	1371	2035
GO:0009058	biosynthetic process	P	0.1637	0.0078	293	359	1140	1733
GO:0005783	endoplasmic reticulum	C	0.2418	0.0127	58	55	1375	2037
GO:0010467	gene expression	P	0.3035	0.0174	141	162	1292	1930
GO:0009536	plastid	C	0.3293	0.0205	6	1	1427	2091
GO:0005739	mitochondrion	C	0.4813	0.0322	113	130	1320	1962
GO:0005576	extracellular region	C	0.6171	0.0443	23	19	1410	2073

P, biological process; F, molecular function; C, cellular component

#Test, number of genes with the designated annotation in test set; #Ref, number of genes with the designated annotation in reference set; #notAnnotTest, number of genes lacking the designated annotation in test set; #notAnnotRef, number of genes lacking the designated annotation in reference set

Table A.4. Enrichment of GO annotations among directional DMGs elevated in haploid males (relative to diploid males) - all P < 0.05

GO-ID	Term	Category	FDR	P-Value	#Test	#Ref	#notAnnotTest	#notAnnotRef
GO:0006807	nitrogen compound metabolic process	P	**0.0060**	8.59E-05	207	405	684	1939
GO:0006139	nucleobase-containing compound metabolic process	P	**0.0060**	8.59E-05	207	405	684	1939
GO:0034641	cellular nitrogen compound metabolic process	P	**0.0060**	8.59E-05	207	405	684	1939
GO:0005488	binding	F	**0.0172**	3.35E-04	626	1497	265	847
GO:0005694	chromosome	C	**0.0172**	4.76E-04	64	98	827	2246
GO:0000166	nucleotide binding	F	**0.0172**	4.93E-04	217	445	674	1899
GO:0044238	primary metabolic process	P	0.0744	0.0025	434	1011	457	1333
GO:0016787	hydrolase activity	F	0.0847	0.0032	211	451	680	1893
GO:0003824	catalytic activity	F	0.0932	0.0040	463	1094	428	1250
GO:0044428	nuclear part	C	0.1054	0.0050	128	257	763	2087
GO:0006259	DNA metabolic process	P	0.1240	0.0071	59	103	832	2241
GO:0090304	nucleic acid metabolic process	P	0.1240	0.0071	59	103	832	2241
GO:0005654	nucleoplasm	C	0.1526	0.0095	91	177	800	2167
GO:0044446	intracellular organelle part	C	0.1621	0.0116	134	280	757	2064
GO:0044422	organelle part	C	0.1621	0.0116	134	280	757	2064
GO:0003676	nucleic acid binding	F	0.1780	0.0157	244	554	647	1790
GO:0031981	nuclear lumen	C	0.1780	0.0170	115	239	776	2105
GO:0031974	membrane-enclosed lumen	C	0.1780	0.0170	115	239	776	2105
GO:0043233	organelle lumen	C	0.1780	0.0170	115	239	776	2105
GO:0070013	intracellular organelle lumen	C	0.1780	0.0170	115	239	776	2105
GO:0008152	metabolic process	P	0.2150	0.0217	522	1279	369	1065
GO:0003682	chromatin binding	F	0.2150	0.0227	18	24	873	2320
GO:0030234	enzyme regulator activity	F	0.2150	0.0237	42	74	849	2270
GO:0003677	DNA binding	F	0.2643	0.0316	94	196	797	2148
GO:0006629	lipid metabolic process	P	0.2643	0.0316	42	76	849	2268
GO:0005634	nucleus	C	0.2702	0.0336	239	554	652	1790
GO:0009056	catabolic process	P	0.2729	0.0353	123	267	768	2077

P, biological process; F, molecular function; C, cellular component

#Test, number of genes with the designated annotation in test set; #Ref, number of genes with the designated annotation in reference set; #notAnnotTest, number of genes lacking the designated annotation in test set; #notAnnotRef, number of genes lacking the designated annotation in reference set

Table A.5. Enrichment of GO annotations among directional DMGs elevated in diploid males (relative to haploid males) - all P < 0.05

GO-ID	Term	Category	FDR	P-Value	#Test	#Ref	#notAnnotTest	#notAnnotRef
GO:0032501	multicellular organismal process	P	0.1739	0.0025	64	291	356	2524
GO:0007275	multicellular organismal development	P	0.1739	0.0025	64	291	356	2524
GO:0032502	developmental process	P	0.1739	0.0031	68	317	352	2498
GO:0009790	embryo development	P	0.1739	0.0033	25	86	395	2729
GO:0040007	growth	P	0.2930	0.0085	16	50	404	2765
GO:0008092	cytoskeletal protein binding	F	0.2930	0.0093	15	46	405	2769
GO:0016049	cell growth	P	0.2930	0.0140	8	18	412	2797
GO:0008361	regulation of cell size	P	0.2930	0.0140	8	18	412	2797
GO:0032535	regulation of cellular component size	P	0.2930	0.0140	8	18	412	2797
GO:0090066	regulation of anatomical structure size	P	0.2930	0.0140	8	18	412	2797
GO:0065008	regulation of biological quality	P	0.3270	0.0172	16	55	404	2760
GO:0030154	cell differentiation	P	0.3357	0.0209	36	163	384	2652
GO:0048869	cellular developmental process	P	0.3357	0.0209	36	163	384	2652
GO:0003779	actin binding	F	0.3622	0.0255	9	25	411	2790
GO:0048856	anatomical structure development	P	0.3622	0.0277	36	167	384	2648
GO:0009653	anatomical structure morphogenesis	P	0.3622	0.0277	36	167	384	2648
GO:0007154	cell communication	P	0.3838	0.0312	17	65	403	2750
GO:0044430	cytoskeletal part	C	0.4449	0.0426	9	28	411	2787
GO:0015630	microtubule cytoskeleton	C	0.4449	0.0426	9	28	411	2787
GO:0005815	microtubule organizing center	C	0.4449	0.0426	9	28	411	2787

P, biological process; F, molecular function; C, cellular component

#Test, number of genes with the designated annotation in test set; #Ref, number of genes with the designated annotation in reference set; #notAnnotTest, number of genes lacking the designated annotation in test set; #notAnnotRef, number of genes lacking the designated annotation in reference set

Table A.6. Differential expression and methylation of putative *S. invicta* dosage compensation[a] orthologs in haploid and diploid males

D. melanogaster **name**	**Gene description[b]**	***S. invicta*** **ortholog**	***S. invicta*** **differential male expression[c]**	***S. invicta*** **elevation of male methylation[d]**
male-specific lethal 1	scaffold for MSL complex assembly	SI2.2.0_02091	diploid > haploid	diploid > haploid
male-specific lethal 2	RING finger protein	SI2.2.0_04411	no data	not significant
male-specific lethal 3	chromodomain protein	SI2.2.0_16160	no data	not significant
males absent on the first	H4K16 acetyltransferase	SI2.2.0_08278	diploid > haploid	not significant
Maleless	RNA/DNA helicase	SI2.2.0_15664	no data	haploid > diploid
Sex lethal	represses msl-2 in females	SI2.2.0_00801	no data	no data
Upstream of N-ras	acts in concert with Sex lethal	SI2.2.0_03090	haploid > diploid	haploid > diploid
Mes-4	H3K36 methyltransferase	SI2.2.0_06678	no data	haploid > diploid
Set2	H3K36 methyltransferase	SI2.2.0_04653	not significant	not significant

[a] Genes associated with dosage compensation here include *D. melanogaster* MSL complex components, regulators of male-specific lethal 2, and H3K36 methyltransferases according to references in footnote b.

[b] Conrad, T. & Akhtar, A. 2012 Dosage compensation in Drosophila melanogaster: epigenetic fine-tuning of chromosome-wide transcription. Nat. Rev. Genet. 13, 123-134.

Gelbart, M.E. & Kuroda, M.I. 2009 Drosophila dosage compensation: a complex voyage to the X chromosome. Development 136, 1399-1410.

Bell, O., Conrad, T., Kind, J., Wirbelauer, C., Akhtar, A. & Schübeler, D. 2008 Transcription-coupled methylation of histone H3 at lysine 36 regulates dosage compensation by enhancing recruitment of the MSL complex in Drosophila melanogaster. Mol. Cell. Biol. 28, 3401-3409.

Wagner, E.J. & Carpenter, P.B. 2012 Understanding the language of Lys36 methylation at histone H3. Nat. Rev. Mol. Cell Biol. 13, 115-126.

[c] Significant expression differences between either pupal or adult haploid and diploid males. No data indicates that probes for these genes were not present on the microarray.

[d] Results of pairwise comparisons between adult haploid and diploid males.

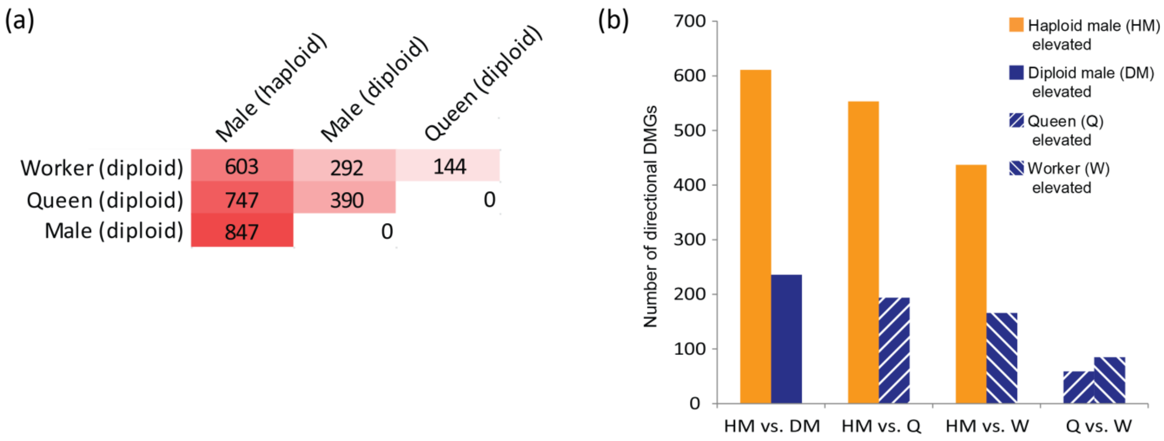


Figure A.1. DNA methylation differs at a 2-fold threshold between haploid and diploid castes in *S. invicta*. This figure depicts analyses comparable to main text figure 1, with differentially methylated genes (DMGs) defined here as those with significant differences according to our generalized linear model, *and* exhibiting a 2-fold difference in DNA methylation level between castes. (a) Number of 2-fold DMGs detected between castes. (b) Number of directional 2-fold DMGs from panel a that exhibit pairwise elevated methylation in haploid (orange) and diploid (blue) castes, respectively.

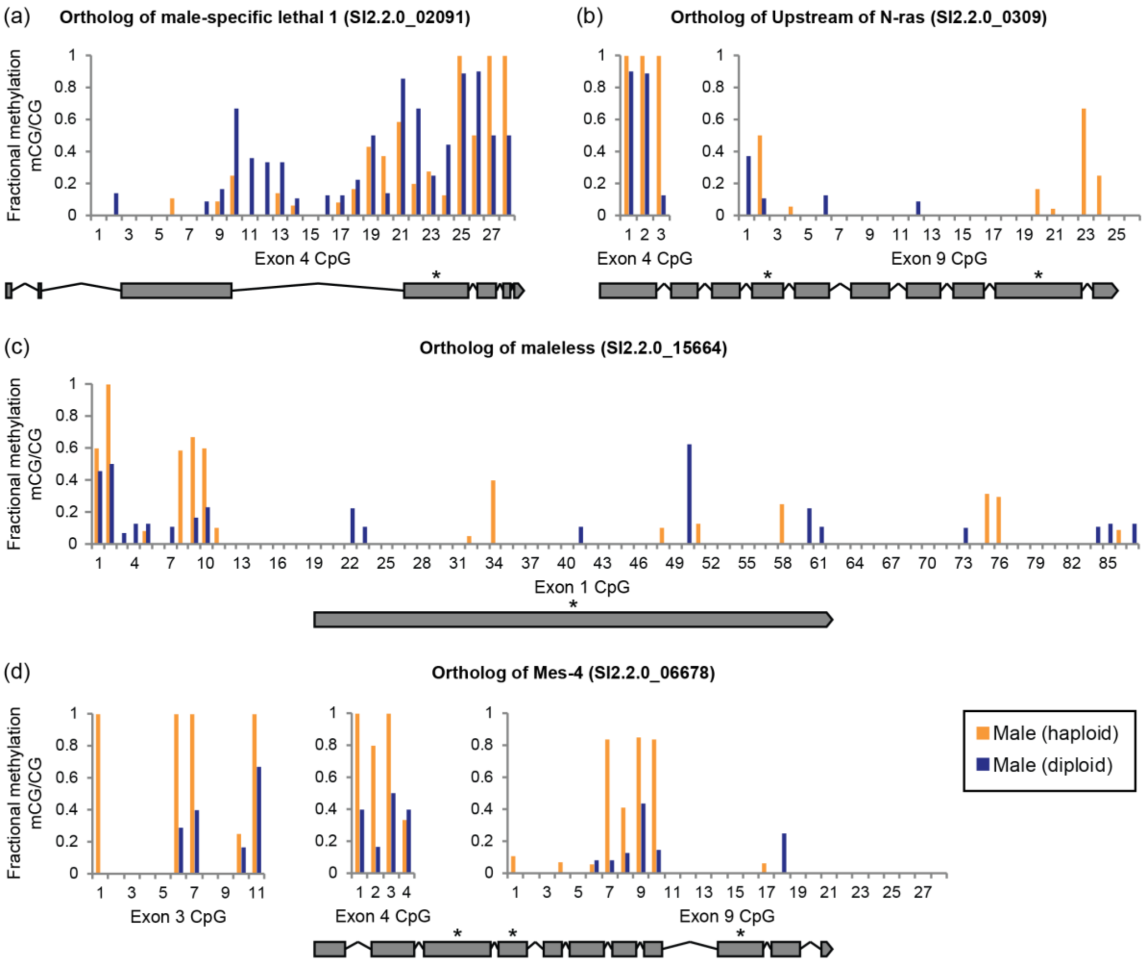


Figure A.2. ***S. invicta*** **DMGs with** ***D. melanogaster*** **orthologs implicated in dosage compensation.** Differential DNA methylation between haploid and diploid males is illustrated for *S. invicta* orthologs of (a) male-specific lethal 1, (b) Upstream of N-ras, (c) maleless, and (d) Mes-4 (genes from table 3 that exhibit differential methylation between haploid and diploid males). Exon 9 of Mes-4 encodes the SET domain (as determined by InterProScan), which is integral to the methylation of H3K36. Each panel shows, at bottom, the associated *S. invicta* gene model. Differentially methylated features (exons and introns) between haploid and diploid males are indicated with an asterisk. Plots show DNA methylation levels for each CpG site in each differentially methylated feature.

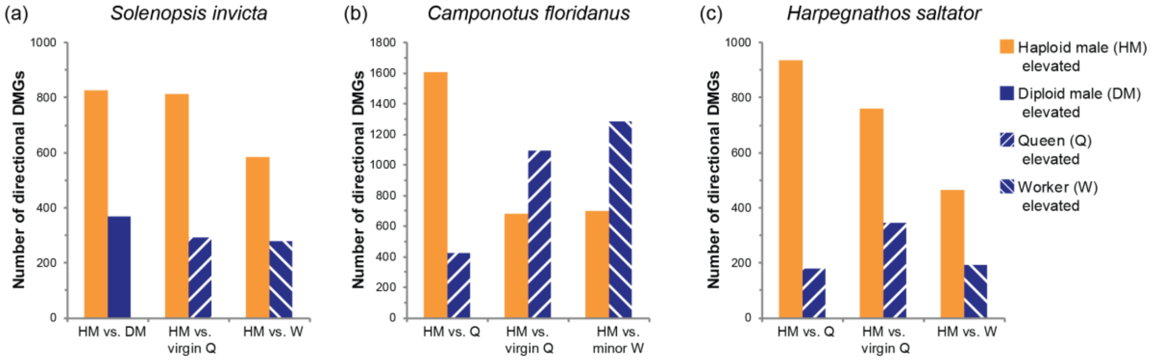


Figure A.3. Directional DMGs frequently exhibit elevated methylation levels in haploid males of multiple ant taxa. The number of directional DMGs that exhibit pairwise elevated methylation in haploid and diploid castes, respectively, from three-way orthologs in the ants (a) *Solenopsis invicta*, (b) *Camponotus floridanus*, and (c) *Harpegnathos saltator*. In *S. invicta* and *H. saltator*, haploid males exhibit elevated methylation more frequently than diploid castes in all comparisons. In *C. floridanus*, haploid males exhibit elevated methylation more frequently than diploid castes in only one of three comparisons. Data from diploid males were not available from *C. floridanus* or *H. saltator*.

APPENDIX B

SUPPLEMENTARY MATERIAL FOR CHAPTER 3

Supplementary Tables and Figures

Table B.1: library read statistics for both RNA- and BS- sequencing libraries.

libID	seq_type	replicate	sample	Raw_pairs	Trimmed pairs	Mapped pairs	Av. coverage
dAFI	BS-seq	1	AF	34,928,624	26,380,282	19,100,078	12.46
dAFII	BS-seq	2	AF	26,614,636	19,074,273	12,994,557	8.81
dAMI	BS-seq	1	AM	60,974,092	30,320,077	27,758,873	16.66
dAMII	BS-seq	2	AM	52,005,005	37,204,129	28,453,831	17.07
dWFI	BS-seq	1	WF	64,489,025	27,277,598	27,240,043	16.34
dWFII	BS-seq	2	WF	58,072,111	27,691,416	26,443,948	15.87
dWMI	BS-seq	1	WM	32,850,028	18,190,383	15,096,203	9.06
dWMII	BS-seq	2	WM	38,681,993	22,044,601	16,649,248	9.99
rAFI	RNA-seq	1	AF	63,640,537	59,033,055	45,690,546	x
rAFII	RNA-seq	2	AF	26,461,888	11,380,495	16,674,255	x
rAFIII	RNA-seq	3	AF	38,111,908	20,071,224	26,064,156	x
rAMI	RNA-seq	1	AM	45,361,857	42,342,312	29,852,380	x
rAMII	RNA-seq	2	AM	54,147,730	31,642,392	33,903,696	x
rAMIII	RNA-seq	3	AM	46,971,271	21,652,370	30,891,506	x
rWFI	RNA-seq	1	WF	42,227,371	39,280,555	30,469,015	x
rWFII	RNA-seq	2	WF	49,998,645	46,773,255	31,453,835	x
rWFIII	RNA-seq	3	WF	27,662,419	13,994,925	15,087,647	x
rWMI	RNA-seq	1	WM	47,492,167	43,803,340	30,233,909	x
rWMII	RNA-seq	2	WM	40,456,443	37,020,907	28,262,964	x
rWMIII	RNA-seq	3	WM	39,925,813	25,111,172	25,310,038	x

Table B.2: Library conversion rates for CpG and CpH sites within the termite genomes, as well as CpG and CpH methylation rates from our unmethylated lambda control.

	CpG	CHG+CHH	CpG	CHH+CHG
sample	genomic	genomic	lambda	lambda
AFI	0.104	0.005	0.004	0.005
AFII	0.101	0.004	0.004	0.004
AMI	0.102	0.004	0.004	0.004
AMII	0.105	0.004	0.004	0.004
WFI	0.105	0.004	0.004	0.004
WFII	0.101	0.004	0.004	0.004
WMI	0.103	0.004	0.004	0.004
WMII	0.106	0.004	0.004	0.004

Table B.3: level of genomic CpG methylation in Znev libraries. mCGs: binomial test-determined “methylated” CpGs; covCGs: total number of CpG’s w/ >4 reads; prop_mCGs: proportion of covered CpGs that are methylated.

lib_or_mapping	mCGs	covCGs	prop_mCGs
ALLCST	1,433,493	11516536	0.119915
AF	1,330,108	11516536	0.112846
AM	1,473,864	11516536	0.11719
WF	1,443,291	11516536	0.115236
WM	1,263,731	11516536	0.11149

Table B.4: level of genomic CpG methylation in Znev libraries.

	ZNEVA				AMELL				CFLOR			
	allCG ct	mCG ct	Prop feature mCG	prop total mCG	allCG ct	mCG ct	Prop feature mCG	prop total mCG	allCG ct`	mCG ct	Prop feature mCG	prop total mCG
genomic	12,872,657	1,546,046	**0.120**	**-**	9,424,047	73,872	**0.008**	**-**	10,295,696	141,700	**0.014**	**-**
frame	3,130,951	1,061,392	**0.339**	**0.668**	4,412,751	67,419	**0.015**	**0.913**	3,066,266	116,331	**0.038**	**0.821**
exons	595,501	346,582	**0.582**	**0.224**	748,622	60,989	**0.081**	**0.826**	1,118,715	92,413	**0.083**	**0.652**
introns	2,532,521	714,348	**0.282**	**0.462**	3,660,706	5,929	**0.002**	**0.080**	1,962,896	23,531	**0.012**	**0.166**
non_genic	8,747,788	449,279	**0.051**	**0.291**	5,011,296	6,190	**0.001**	**0.084**	7,229,430	21,082	**0.003**	**0.149**

Table B.5: Differential methylation calls between libraries. For caste and sex the number of differentially methylated genes are presented, as well as the representative number of hypermethylated genes in both compared samples. "Combined framework": represents numbers of differentially methylated genes between caste and sex while controlling for the alternative. "Pairwise tests": number of differentially methylated genes when comparing each representative pair separately for caste and sex. Also given for pairwise tests is number of differentially methylated genes that show consistent directional differential methylation between both pairs of a given comparison (caste or sex):

	test		**totals**	hyper-caste	
				A	W
combined framework	caste		**2,611**	2,515	96
				F	M
	sex		**209**	114	95
				A	W
pairwise tests	caste	AF.WF	**4,380**	4,243	137
		AM.WM	**1,733**	1,448	285
		shared	**1,110**	1,098	12
				F	M
	sex	AF.AM	**593**	472	121
		WF.WM	**1,393**	357	1,036
		shared	**36**	25	11

Table B.6: statistics of genic localization for DMRs and DMCs for caste- and sex-significant DMRs/DMCs.

	cmp	feature	proportion of DMRs	significant DMR count	tested regions
DMRs	caste	intron	0.441	6,155	83,062
		exon	0.380	5,309	58,534
		3prox	0.108	1,512	18,936
		5prox	0.071	988	13,425
		total		13,964	173,957
	sex	intron	0.435	346	83,866
		exon	0.345	274	58,871
		3prox	0.130	104	19,094
		5prox	0.089	71	13,579
		total		795	175,410
DMCs	caste	intron	0.392	2,822	369,785
		exon	0.422	3,042	339,058
		3prox	0.111	800	91,326
		5prox	0.075	540	56,121
		total		7,204	856,289
	sex	intron	0.411	556	372,033
		exon	0.347	470	339,993
		3prox	0.142	192	91,864
		5prox	0.100	135	56,450
		total		1,353	860,340

Table B.7: Gene ontology enrichment for genes featuring significant DNA methylation, and those featuring no DNA methylation in the *Z. nevadensis* genome.

Term	Category	FDR	Fold enrichment	GO-ID
Methylated genes				
ATP binding	F	9.88E-30	5.74	GO:0005524
protein phosphorylation	P	4.01E-07	2.63	GO:0006468
DNA repair	P	4.05E-07	13.91	GO:0006281
macromolecular complex subunit organization	P	8.63E-07	2.42	GO:0043933
zinc ion binding	F	9.35E-07	1.73	GO:0008270
microtubule organizing center	C	3.22E-06	21.87	GO:0005815
protein serine/threonine kinase activity	F	3.24E-06	2.86	GO:0004674
spliceosomal complex	C	2.21E-05	8.23	GO:0005681
nucleoplasm part	C	3.10E-05	4.72	GO:0044451
ATP-dependent helicase activity	F	3.31E-05	16.42	GO:0008026
translation factor activity, RNA binding	F	6.25E-05	10.33	GO:0008135
purine ribonucleoside triphosphate catabolic process	P	6.25E-05	10.33	GO:0009207
histone modification	P	9.97E-05	7.60	GO:0016570
RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	P	1.14E-04	4.34	GO:0000377
single-organism carbohydrate metabolic process	P	1.14E-04	4.37	GO:0044723
structural constituent of ribosome	F	2.33E-04	5.12	GO:0003735
spindle	C	3.65E-04	13.11	GO:0005819
chromosome	C	3.94E-04	2.47	GO:0005694
Golgi vesicle transport	P	5.43E-04	12.64	GO:0048193
oxidoreductase activity, acting on the CH-CH group of donors	F	8.18E-04	12.16	GO:0016627
oogenesis	P	8.98E-04	4.70	GO:0048477
nuclear envelope	C	1.25E-03	11.69	GO:0005635
motor activity	F	1.27E-03	8.32	GO:0003774
regulation of hydrolase activity	P	1.75E-03	4.46	GO:0051336
transferase activity, transferring acyl groups other than amino-acyl groups	F	2.23E-03	3.44	GO:0016747
mitotic nuclear division	P	2.43E-03	5.94	GO:0007067
microtubule-based movement	P	2.52E-03	13.34	GO:0007018
tRNA aminoacylation for protein translation	P	2.70E-03	10.52	GO:0006418
mitotic M phase	P	3.72E-03	12.87	GO:0000087
GTPase regulator activity	F	3.82E-03	7.61	GO:0030695
isomerase activity	F	4.18E-03	4.80	GO:0016853
ubiquitin-like protein transferase activity	F	4.18E-03	4.80	GO:0019787
regulation of Rho protein signal transduction	P	4.71E-03	4.17	GO:0035023
ribosomal subunit	C	4.84E-03	5.54	GO:0044391
monocarboxylic acid metabolic process	P	5.11E-03	5.70	GO:0032787

microtubule cytoskeleton organization	P	5.38E-03	2.91	GO:0000226
male gamete generation	P	5.40E-03	12.39	GO:0048232
endoplasmic reticulum membrane	C	5.40E-03	12.39	GO:0005789
microtubule associated complex	C	6.08E-03	9.58	GO:0005875
aminoacyl-tRNA ligase activity	F	6.34E-03	9.81	GO:0004812
Unmethylated genes				
structural constituent of cuticle	F	5.56E-25	20.52	GO:0042302
sequence-specific DNA binding	F	5.30E-11	3.18	GO:0043565
odorant binding	F	8.27E-10	27.54	GO:0005549
sequence-specific DNA binding transcription factor activity	F	2.74E-08	2.35	GO:0003700
chitin binding	F	1.87E-07	5.89	GO:0008061
heme binding	F	3.27E-06	3.08	GO:0020037
neuropeptide receptor activity	F	4.13E-06	18.78	GO:0008188
oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	F	5.03E-06	3.01	GO:0016705
chitin metabolic process	P	7.19E-06	4.21	GO:0006030
development of primary sexual characteristics	P	1.55E-05	3.83	GO:0045137
electron carrier activity	F	1.72E-04	2.62	GO:0009055
nucleosome	C	2.89E-04	5.91	GO:0000786
hormone activity	F	4.15E-04	9.52	GO:0005179
regulation of transcription, DNA-templated	P	4.85E-04	1.53	GO:0006355
integral component of membrane	C	6.85E-04	1.49	GO:0016021
extracellular region	C	7.18E-04	2.12	GO:0005576
nucleosome assembly	P	2.17E-03	4.33	GO:0006334
flavin adenine dinucleotide binding	F	2.41E-03	3.08	GO:0050660
cell fate specification	P	3.59E-03	3.39	GO:0001708
cullin-RING ubiquitin ligase complex	C	5.85E-03	2.89	GO:0031461
G-protein coupled amine receptor activity	F	6.23E-03	8.64	GO:0008227
carboxylic ester hydrolase activity	F	1.09E-02	3.20	GO:0052689
G-protein coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger	P	1.25E-02	6.91	GO:0007187
positive regulation of sodium ion transport	P	1.80E-02	21.57	GO:0010765
Wnt signaling pathway, calcium modulating pathway	P	1.92E-02	7.56	GO:0007223
metalloexopeptidase activity	F	2.09E-02	3.96	GO:0008235
DNA integration	P	2.20E-02	5.76	GO:0015074
sodium channel activity	F	2.73E-02	3.46	GO:0005272
central nervous system development	P	3.33E-02	1.83	GO:0007417
neural tube development	P	3.54E-02	6.05	GO:0021915
axon extension	P	3.68E-02	4.94	GO:0048675
extracellular-glutamate-gated ion channel activity	F	4.42E-02	10.79	GO:0005234
enteroendocrine cell differentiation	P	4.42E-02	10.79	GO:0035883

cGMP biosynthetic process	P	4.42E-02	10.79	GO:0006182
regulation of muscle organ development	P	4.42E-02	10.79	GO:0048634
guanylate cyclase activity	F	4.42E-02	10.79	GO:0004383
specification of segmental identity, head	P	4.42E-02	10.79	GO:0007380
endocrine pancreas development	P	4.42E-02	10.79	GO:0031018
morphogenesis of a branching epithelium	P	4.62E-02	2.54	GO:0061138
negative regulation of multicellular organismal process	P	4.79E-02	3.60	GO:0051241

Table B.8: GO terms associated with the highest and lowest two methylation deciles, relative to all other methylated genes:

Term	Category	FDR	fold enrichment	GO-ID
		Highest deciles		
chromatin modification	P	2.43E-03	3.22	GO:0016568
nucleic acid binding	F	2.73E-03	1.67	GO:0003676
transition metal ion binding	F	3.85E-03	1.67	GO:0046914
chromatin organization	P	7.88E-03	2.68	GO:0006325
chromosome organization	P	1.54E-02	2.19	GO:0051276
metal ion binding	F	1.54E-02	1.52	GO:0046872
macromolecule methylation	P	1.54E-02	4.55	GO:0043414
histone modification	P	1.54E-02	3.07	GO:0016570
covalent chromatin modification	P	1.54E-02	3.07	GO:0016569
zinc ion binding	F	1.54E-02	1.65	GO:0008270
nucleic acid metabolic process	P	1.59E-02	1.52	GO:0090304
methylation	P	1.59E-02	4.11	GO:0032259
nucleus	C	1.65E-02	1.59	GO:0005634
chromosome	C	1.94E-02	2.22	GO:0005694
macromolecular complex subunit organization	P	2.51E-02	1.88	GO:0043933
protein-lysine N-methyltransferase activity	F	2.51E-02	6.66	GO:0016279
lysine N-methyltransferase activity	F	2.51E-02	6.66	GO:0016278
S-adenosylmethionine-dependent methyltransferase activity	F	2.51E-02	4.05	GO:0008757
regulation of gene expression	P	3.67E-02	1.68	GO:0010468
transferase activity, transferring one-carbon groups	F	4.23E-02	2.56	GO:0016741
histone methylation	P	4.23E-02	5.18	GO:0016571
cell fate commitment	P	4.31E-02	2.92	GO:0045165
cation binding	F	4.62E-02	1.44	GO:0043169
multi-organism process	P	4.98E-02	1.79	GO:0051704
DNA methylation or demethylation	P	4.98E-02	41.20	GO:0044728
		Lowest deciles		
signaling receptor activity	F	2.67E-05	3.28	GO:0038023
G-protein coupled receptor activity	F	2.73E-05	4.92	GO:0004930
receptor activity	F	2.73E-05	2.93	GO:0004872
transmembrane signaling receptor activity	F	8.47E-05	3.35	GO:0004888
integral component of membrane	C	9.14E-05	1.86	GO:0016021
intrinsic component of membrane	C	1.88E-04	1.68	GO:0031224
membrane	C	3.95E-04	1.47	GO:0016020
amino acid transmembrane transporter activity	F	1.26E-03	15.06	GO:0015171
heme binding	F	3.02E-03	3.54	GO:0020037
membrane part	C	3.22E-03	1.50	GO:0044425
molecular transducer activity	F	3.52E-03	1.90	GO:0060089

tetrapyrrole binding	F	3.52E-03	3.45	GO:0046906
transporter activity	F	3.59E-03	1.70	GO:0005215
signal transducer activity	F	3.81E-03	1.96	GO:0004871
carboxylic acid transmembrane transporter activity	F	4.72E-03	7.44	GO:0046943
organic anion transmembrane transporter activity	F	4.72E-03	7.44	GO:0008514
organic acid transmembrane transporter activity	F	4.72E-03	7.44	GO:0005342
transmembrane transporter activity	F	8.02E-03	1.71	GO:0022857
G-protein coupled receptor signaling pathway	P	9.91E-03	2.87	GO:0007186
circadian behavior	P	1.03E-02	20.03	GO:0048512
electron carrier activity	F	1.92E-02	2.99	GO:0009055
Wnt signaling pathway, calcium modulating pathway	P	2.13E-02	33.23	GO:0007223
circadian sleep/wake cycle	P	4.14E-02	16.67	GO:0042745
transmembrane transport	P	4.88E-02	1.73	GO:0055085
oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	F	6.66E-02	2.63	GO:0016705

TABLE B.9: Terms associated with genes methylated in *Z. nevadensis* that are not methylated in the other species examined in this study:

Term	Cat	FDR	fold enrich	GO-ID
calcium ion binding	F	8.65E-07	4.29	GO:0005509
sequence-specific DNA binding	F	3.11E-06	4.82	GO:0043565
integral component of membrane	C	3.48E-05	2.15	GO:0016021
cell projection	C	4.59E-05	3.07	GO:0042995
extracellular region	C	2.10E-04	4.02	GO:0005576
nucleic acid binding transcription factor activity	F	3.04E-04	1.99	GO:0001071
G-protein coupled receptor signaling pathway	P	3.74E-04	6.57	GO:0007186
rhythmic process	P	5.73E-04	13.55	GO:0048511
open tracheal system development	P	6.05E-04	5.12	GO:0007424
eye development	P	6.05E-04	2.75	GO:0001654
locomotory behavior	P	2.15E-03	4.49	GO:0007626
epithelial cell migration	P	2.15E-03	5.75	GO:0010631
serine-type endopeptidase activity	F	3.08E-03	6.01	GO:0004252
transmembrane signaling receptor activity	F	3.09E-03	4.55	GO:0004888
membrane	C	4.81E-03	1.34	GO:0016020
proteinaceous extracellular matrix	C	4.81E-03	11.09	GO:0005578
sensory perception of chemical stimulus	P	5.46E-03	14.79	GO:0007606
axon choice point recognition	P	5.46E-03	14.79	GO:0016198
multi-organism behavior	P	7.10E-03	4.27	GO:0051705
sexual reproduction	P	9.04E-03	1.95	GO:0019953
cell fate commitment	P	1.16E-02	3.49	GO:0045165
homophilic cell adhesion via plasma membrane adhesion molecules	P	1.42E-02	9.86	GO:0007156
growth factor activity	F	1.68E-02	22.16	GO:0008083
formation of primary germ layer	P	1.79E-02	12.94	GO:0001704
imaginal disc pattern formation	P	2.02E-02	5.08	GO:0007447
multicellular organismal reproductive process	P	2.25E-02	1.87	GO:0048609
alpha-amino acid metabolic process	P	2.37E-02	2.62	GO:1901605
single organism reproductive process	P	2.64E-02	1.85	GO:0044702
synapse	C	2.88E-02	5.28	GO:0045202
regulation of transcription, DNA-templated	P	2.92E-02	1.54	GO:0006355
sensory organ morphogenesis	P	3.23E-02	2.29	GO:0090596
cell periphery	C	3.31E-02	2.03	GO:0071944
adult behavior	P	3.41E-02	4.03	GO:0030534
peptidase inhibitor activity	F	3.77E-02	5.55	GO:0030414
dorsal/ventral pattern formation	P	3.77E-02	5.55	GO:0009953
single organismal cell-cell adhesion	P	3.77E-02	5.55	GO:0016337
imaginal disc morphogenesis	P	3.77E-02	2.24	GO:0007560
cell-cell signaling	P	3.92E-02	2.50	GO:0007267

regulation of cell morphogenesis	P	4.41E-02	4.62	GO:0022604
plasma membrane part	C	4.70E-02	2.18	GO:0044459
enzyme linked receptor protein signaling pathway	P	4.70E-02	2.27	GO:0007167
protein phosphorylation	P	4.72E-02	1.69	GO:0006468
photoreceptor cell development	P	4.72E-02	3.43	GO:0042461
epithelial structure maintenance	P	4.72E-02	11.09	GO:0010669
serine-type endopeptidase inhibitor activity	F	4.72E-02	11.09	GO:0004867
mesoderm morphogenesis	P	4.72E-02	11.09	GO:0048332
response to wounding	P	4.72E-02	11.09	GO:0009611
embryonic heart tube morphogenesis	P	4.72E-02	11.09	GO:0003143
regulation of multicellular organismal process	P	4.92E-02	2.20	GO:0051239

Table B.10. Terms significantly enriched among differentially methylated genes that are methylated in *Z. nevadensis* but not *C. floridanus* or *A. mellifera*, relative to differentially methylated genes that are methylated in all species:

Term	Cat	FDR	fold enrich	GO-ID
multicellular organismal process	P	3.01E-04	1.89	GO:0032501
single-multicellular organism process	P	3.01E-04	1.94	GO:0044707
anatomical structure morphogenesis	P	3.71E-04	2.49	GO:0009653
developmental process	P	1.02E-03	1.82	GO:0032502
single-organism developmental process	P	1.02E-03	1.85	GO:0044767
sensory perception	P	1.46E-03	37.79	GO:0007600
cellular developmental process	P	3.69E-03	2.19	GO:0048869
anatomical structure development	P	3.69E-03	1.84	GO:0048856
signal transducer activity	F	7.67E-03	3.82	GO:0004871
multicellular organismal development	P	7.74E-03	1.79	GO:0007275
system development	P	7.74E-03	1.92	GO:0048731
sequence-specific DNA binding	F	7.74E-03	7.20	GO:0043565
biological adhesion	P	9.38E-03	5.34	GO:0022610
cell adhesion	P	9.38E-03	5.34	GO:0007155
organ morphogenesis	P	1.05E-02	2.92	GO:0009887
cell development	P	1.12E-02	2.44	GO:0048468
protein dimerization activity	F	1.38E-02	8.40	GO:0046983
molecular transducer activity	F	1.45E-02	3.08	GO:0060089
organ development	P	1.45E-02	2.10	GO:0048513
cell differentiation	P	1.63E-02	2.07	GO:0030154
regionalization	P	1.63E-02	3.76	GO:0003002
system process	P	1.79E-02	4.20	GO:0003008
tissue development	P	2.07E-02	2.32	GO:0009888
extracellular region	C	2.07E-02	7.00	GO:0005576
dorsal/ventral pattern formation	P	2.07E-02	11.20	GO:0009953
imaginal disc pattern formation	P	2.07E-02	11.20	GO:0007447
cell surface receptor signaling pathway	P	2.72E-02	2.56	GO:0007166
proteinaceous extracellular matrix	C	3.00E-02	5.88	GO:0005578
nucleic acid binding transcription factor activity	F	3.07E-02	2.35	GO:0001071
signaling	P	3.28E-02	1.62	GO:0023052
calcium ion binding	F	3.44E-02	3.92	GO:0005509
neurological system process	P	3.44E-02	4.20	GO:0050877
respiratory system development	P	3.44E-02	5.13	GO:0060541
cell morphogenesis	P	3.44E-02	2.95	GO:0000902
pattern specification process	P	3.44E-02	2.95	GO:0007389
extracellular region part	C	3.75E-02	8.40	GO:0044421
cellular component morphogenesis	P	4.28E-02	2.67	GO:0032989
epithelium development	P	4.55E-02	2.32	GO:0060429

Table B.11: terms enriched for DMR-containing genes relative to methylated genes that do not contain any significantly-differing DMRs:

Term	Cat	FDR	fold enrich	GO-ID
protein binding	F	5.04E-10	1.30	GO:0005515
ATP binding	F	1.51E-07	1.62	GO:0005524
binding	F	9.33E-07	1.11	GO:0005488
adenyl ribonucleotide binding	F	1.28E-06	1.50	GO:0032559
adenyl nucleotide binding	F	1.28E-06	1.50	GO:0030554
nucleoside phosphate binding	F	1.28E-06	1.40	GO:1901265
nucleotide binding	F	1.28E-06	1.40	GO:0000166
purine nucleoside binding	F	1.28E-06	1.47	GO:0001883
purine ribonucleoside binding	F	1.28E-06	1.47	GO:0032550
ribonucleoside binding	F	1.28E-06	1.47	GO:0032549
purine ribonucleoside triphosphate binding	F	1.28E-06	1.47	GO:0035639
nucleoside binding	F	1.35E-06	1.47	GO:0001882
small molecule binding	F	2.37E-06	1.38	GO:0036094
anion binding	F	2.56E-06	1.41	GO:0043168
purine ribonucleotide binding	F	4.25E-06	1.41	GO:0032555
ribonucleotide binding	F	4.25E-06	1.41	GO:0032553
purine nucleotide binding	F	4.25E-06	1.41	GO:0017076
carbohydrate derivative binding	F	1.04E-05	1.38	GO:0097367
single-organism cellular process	P	1.11E-05	1.17	GO:0044763
cellular process	P	7.52E-05	1.10	GO:0009987
single-organism process	P	6.77E-04	1.12	GO:0044699
signaling	P	9.07E-04	1.30	GO:0023052
heterocyclic compound binding	F	1.11E-03	1.19	GO:1901363
organic cyclic compound binding	F	1.14E-03	1.18	GO:0097159
signal transduction	P	1.24E-03	1.32	GO:0007165
ion binding	F	1.24E-03	1.17	GO:0043167
cell communication	P	1.80E-03	1.29	GO:0007154
organelle organization	P	1.91E-03	1.36	GO:0006996
phosphate-containing compound metabolic process	P	1.91E-03	1.33	GO:0006796
protein kinase activity	F	2.65E-03	1.62	GO:0004672
single organism signaling	P	3.00E-03	1.29	GO:0044700
ATPase activity	F	3.48E-03	1.72	GO:0016887
cellular response to stimulus	P	3.78E-03	1.26	GO:0051716
phosphorus metabolic process	P	4.35E-03	1.31	GO:0006793
response to stimulus	P	4.96E-03	1.20	GO:0050896
protein serine/threonine kinase activity	F	5.13E-03	1.63	GO:0004674
multicellular organismal process	P	5.58E-03	1.22	GO:0032501
regulation of cellular process	P	5.58E-03	1.19	GO:0050794

cellular component organization	P	5.73E-03	1.25	GO:0016043
regulation of biological process	P	6.20E-03	1.17	GO:0050789
regulation of cell projection organization	P	6.21E-03	6.79	GO:0031344
Rho protein signal transduction	P	6.52E-03	2.29	GO:0007266
small GTPase binding	F	6.91E-03	19.02	GO:0031267
GTPase binding	F	7.38E-03	10.87	GO:0051020
actin filament-based process	P	8.11E-03	2.07	GO:0030029
calcium ion binding	F	8.36E-03	2.01	GO:0005509
single-multicellular organism process	P	8.45E-03	1.23	GO:0044707
protein phosphorylation	P	8.45E-03	1.50	GO:0006468
phosphotransferase activity, alcohol group as acceptor	F	8.49E-03	1.49	GO:0016773
nucleoside-triphosphatase activity	F	9.87E-03	1.39	GO:0017111
Ras protein signal transduction	P	1.01E-02	2.07	GO:0007265
regulation of Rho protein signal transduction	P	1.11E-02	2.26	GO:0035023
cell part	C	1.20E-02	1.08	GO:0044464
Ras GTPase binding	F	1.20E-02	17.66	GO:0017016
intracellular signal transduction	P	1.22E-02	1.42	GO:0035556
hydrolase activity, acting on acid anhydrides	F	1.27E-02	1.37	GO:0016817
actin cytoskeleton organization	P	1.30E-02	2.02	GO:0030036
phosphorylation	P	1.32E-02	1.45	GO:0016310
anatomical structure development	P	1.34E-02	1.23	GO:0048856
cellular component organization or biogenesis	P	1.50E-02	1.21	GO:0071840
hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	F	1.71E-02	1.36	GO:0016818
pyrophosphatase activity	F	1.73E-02	1.36	GO:0016462
regulation of intracellular signal transduction	P	2.21E-02	1.76	GO:1902531
regulation of Ras protein signal transduction	P	2.21E-02	2.07	GO:0046578
single-organism organelle organization	P	2.26E-02	1.38	GO:1902589
regulation of response to stimulus	P	2.33E-02	1.50	GO:0048583
multicellular organismal development	P	2.51E-02	1.22	GO:0007275
regulation of cellular component organization	P	2.60E-02	1.71	GO:0051128
organophosphate catabolic process	P	3.05E-02	2.00	GO:0046434
single-organism developmental process	P	3.05E-02	1.20	GO:0044767
biological regulation	P	3.07E-02	1.14	GO:0065007
cytoskeleton organization	P	3.07E-02	1.49	GO:0007010
organelle part	C	3.08E-02	1.18	GO:0044422
motor activity	F	3.09E-02	2.30	GO:0003774
system development	P	3.47E-02	1.25	GO:0048731
intracellular organelle part	C	3.51E-02	1.19	GO:0044446
molecular function regulator	F	3.66E-02	1.52	GO:0098772
double-stranded RNA binding	F	3.98E-02	14.95	GO:0003725
chromosome	C	3.98E-02	1.56	GO:0005694
hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of	F	3.99E-02	2.32	GO:0016820

substances				
transferase activity, transferring phosphorus-containing groups	F	3.99E-02	1.32	GO:0016772
plasma membrane	C	4.13E-02	1.69	GO:0005886
cellular protein modification process	P	4.26E-02	1.24	GO:0006464
protein modification process	P	4.26E-02	1.24	GO:0036211
regulation of signal transduction	P	4.41E-02	1.52	GO:0009966
nucleus	C	4.41E-02	1.22	GO:0005634
embryonic pattern specification	P	4.41E-02	2.50	GO:0009880
plasma membrane part	C	4.41E-02	1.75	GO:0044459
anatomical structure morphogenesis	P	4.66E-02	1.28	GO:0009653
regulation of small GTPase mediated signal transduction	P	4.77E-02	1.85	GO:0051056

TABLE B.12: Gene Ontology terms enriched among genes containing DMRs that differ significantly between castes and sexes relative to all DMR-containing genes:

CASTE

Term	Cat	FDR	FE	GO-ID
protein binding	F	3.60E-06	1.25	GO:0005515
purine ribonucleotide binding	F	2.21E-04	1.39	GO:0032555
ATP binding	F	3.16E-04	1.48	GO:0005524
protein complex	C	5.60E-04	1.40	GO:0043234
phosphate-containing compound metabolic process	P	1.33E-03	1.35	GO:0006796
intracellular membrane-bounded organelle	C	1.33E-03	1.18	GO:0043231
small molecule metabolic process	P	4.77E-03	1.35	GO:0044281
organonitrogen compound catabolic process	P	4.77E-03	1.80	GO:1901565
cell cycle phase	P	7.06E-03	2.41	GO:0022403
cellular amino acid metabolic process	P	9.05E-03	1.85	GO:0006520
single-organism catabolic process	P	1.09E-02	1.93	GO:0044712
cytoskeleton organization	P	1.68E-02	1.60	GO:0007010
cytoskeleton	C	2.22E-02	1.43	GO:0005856
nucleoplasm part	C	2.69E-02	1.94	GO:0044451
cytoplasmic part	C	3.24E-02	1.23	GO:0044444
aromatic compound catabolic process	P	4.07E-02	1.50	GO:0019439
organophosphate catabolic process	P	4.95E-02	1.57	GO:0046434

SEX

Term	Cat	FDR	FE	GO-ID
response to stimulus	P	5.25E-03	1.89	GO:0050896
nucleotide binding	F	1.51E-02	1.36	GO:0000166
regulation of biological process	P	2.41E-02	2.23	GO:0050789
cell communication	P	3.13E-02	1.33	GO:0007154
nucleoside-triphosphatase activity	F	4.88E-02	1.50	GO:0017111
anion binding	F	4.88E-02	1.33	GO:0043168
single organism signaling	P	4.88E-02	1.31	GO:0044700

TABLE B.13: Gene ontology terms enriched for sex- or caste-specific differential expression.

	Category	FDR
SEX		
DNA binding	F	1.03E-11
zinc ion binding	F	2.96E-06
protein complex	C	3.21E-05
chromatin assembly or disassembly	P	3.99E-05
DNA conformation change	P	3.99E-05
nucleosome assembly	P	6.99E-05
nucleosome	C	6.99E-05
RNA processing	P	8.34E-05
mitotic nuclear division	P	2.56E-03
DNA-dependent DNA replication	P	8.16E-03
ubiquitin-dependent protein catabolic process	P	8.20E-03
cell division	P	1.10E-02
meiotic nuclear division	P	1.10E-02
cell cycle phase	P	1.39E-02
regulation of cellular metabolic process	P	2.13E-02
regulation of primary metabolic process	P	2.13E-02
transferase complex	C	2.38E-02
DNA repair	P	3.03E-02
DNA helicase activity	F	3.29E-02
meiotic cell cycle process	P	3.29E-02
histone acetyltransferase complex	C	3.29E-02
chromosomal region	C	3.50E-02
chromatin modification	P	3.50E-02
histone modification	P	3.84E-02
regulation of gene expression	P	3.86E-02
CASTE		
oxidoreductase activity	F	1.99E-07
membrane	C	3.81E-04
small molecule metabolic process	P	4.68E-04
organonitrogen compound biosynthetic process	P	7.75E-04
transmembrane transporter activity	F	2.94E-03
single-organism biosynthetic process	P	4.80E-03
single-organism carbohydrate metabolic process	P	5.38E-03
cytoplasmic part	C	5.90E-03
substrate-specific transporter activity	F	8.22E-03
transmembrane transport	P	1.19E-02
integral component of plasma membrane	C	2.02E-02
ribose phosphate metabolic process	P	2.22E-02

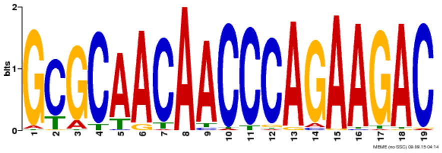
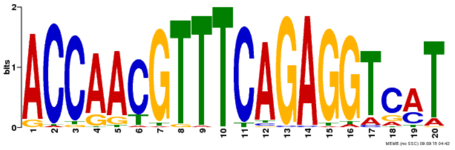
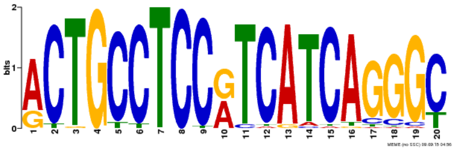
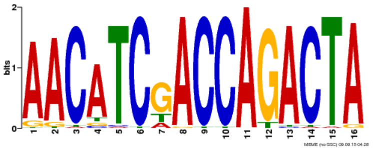
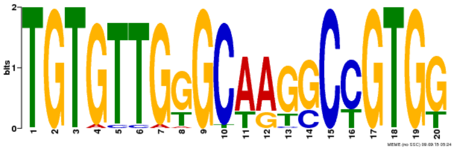
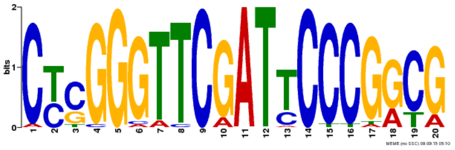
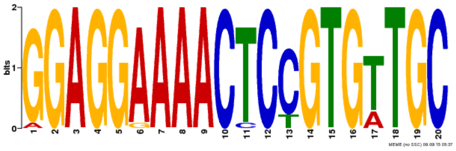
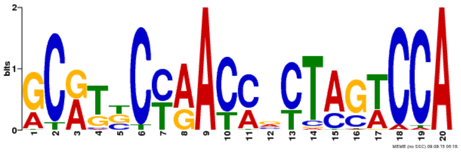
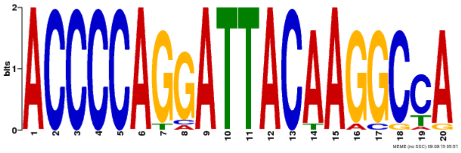
cellular respiration	P	3.24E-02
alpha-amino acid metabolic process	P	3.24E-02
monovalent inorganic cation transmembrane transporter activity	F	3.37E-02
carboxylic acid biosynthetic process	P	4.44E-02
purine ribonucleotide metabolic process	P	4.55E-02

Table B.14: results from DMR-associated TFBS analyses. AME qvales: qvalue from AME program comparing DMRs to non-DMR CpG-centered nearby sequences for caste and sex DMRs. Clover siglvls: level of significance when using the CLOVER program to compare DMRs to control sequences (1: significant test when comparing DMRs to non-DMR nearby mCpG-centered sequences, 2: significant as in 1, but also when comparing to the sequences from all methylated introns on associated genes); Type: bolded values indicate confidently enriched motifs: consistent significance with both programs and the existence of a *Z. nevadensis* ortholog to the given TF; CvS: whether the given TF was significantly enriched in sex- or caste- DMRs relative to the other.

TF	AME CST q	AME SEX q	Clover cst siglvl	Clover sex siglvl	Znev gnID	type	CvS comp	longname
p120	2.09E-06		1		Znev_00683	caste		Myb-interacting protein 120
Eip74EF	7.59E-06		2		Znev_00833	**caste**	**caste**	Ecdysone-induced protein 74EF
fkh	1.00E-05		2		Znev_13477	**caste**	**caste**	fork head
Ubx	3.93E-05		2		Znev_15380	**caste**	**caste**	ultrabithorax
bab1	7.15E-05		2		Znev_03179	**caste**	**caste**	bric a brac
en	5.28E-03		2		Znev_15553	**caste**		engrailed
nub		8.55E-03	2	2	Znev_14256	**sex**		nubbin
z		4.55E-02	1	2	Znev_02821	**sex**	**sex**	zeste
exd	7.80E-03		1		Znev_12397	caste		extradenticle
hb	3.70E-02		1		Znev_01840	caste	caste	hunchback
hkb	6.42E-03		1		NA	caste		hucklebein
zen	2.51E-03				NA	caste		zerknullt
Aef1	1.76E-03				NA	caste		Adult enhancer factor 1
gsb	3.52E-02				NA	caste		gooseberry
tll			2		Znev_12982	caste		tailless
SuH			2	1	Znev_04163	caste		supressor of hairless
gt			2		NA	caste		giant
srp			2		Znev_02318	caste		serpent
usp			2		Znev_11534	caste		ultraspiracle
br			1	2	Znev_09723	sex		broad
Hsf			1	2	Znev_08108	sex		heat shock factor
vvl			1	1	Znev_11549	caste		ventral veins lacking
Dfd			1		Znev_05733	caste	caste	Deformed
ap			1		Znev_18686	caste		apterous
ems			1		Znev_10939	caste	caste	empty spiracles
Hr46			1		Znev_14707	caste		Hormone receptor-like in 46
vnd				1	Znev_00117	sex		ventral nervous system defective

ttk	1	Znev_15196	sex	sex	tram track
grh	1	Znev_13872	sex		grainy head

Table B.15: top 10 significantly enriched motifs as determined by MEME *de novo* motif discovery, alongside any significantly-similar (q-value < 0.25) miRNA or TFBS sequence motif (Similarity hits).

Motif logo	width	evalue	Similarity hits
	19	3.50E-260	Eip74EF
	20	1.10E-227	Zne-mir-87-2-3p, Zne-mir-87-2-3p, Zne-mir-87-3-3p, Zne-mir-6012-5p, Zne-mir-282-3p, Zne-mir-316-3p, Zne-mir-9d-3p
	20	6.00E-200	dme-miR-313-5p, Zne-mir-34-3p
	16	3.60E-141	NA
	20	4.00E-103	NA
	20	1.80E-98	dme-miR-4951-5p, dme-miR-4952-3p
	20	1.90E-79	dl
	20	5.00E-61	bcd
	20	5.60E-59	Zne-mir-184-5p

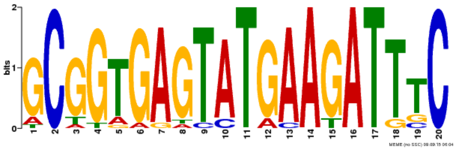
	20	5.90E-48	dme-miR-1003-3p

Table B.16: results of DMR miRNA homology tests. For each Znev miRNA showing homology to at least one DMR (89/190) Produced sing the AME and FIMO tools from the MEME suite. Counts and proportion (of total counts) represent number of DMRs showing significant (FDR < 0.1) homology to the given miRNA sequence for Caste and Sex differing DMRs. Qvalues represent results of testing whether DMR sequences show higher representation of the given mature miRNA than surrounding methylated region (methylated regions up and down-stream of DMR showing no differential methylation).

	CASTE				SEX			
miRNA	positive hits	negative hits	fold difference	AME FDR	positive hits	negative hits	fold difference	AME FDR
Zne-bantam-3p	0	3	0.470	ns	0	1	0.921	ns
Zne-bantam-5p	7	6	2.192	1.02E-07	14	4	6.448	0.03721
Zne-let-7-3p	0	0	NA	ns	0	0	NA	ns
Zne-let-7-5p	0	0	NA	ns	0	0	NA	ns
Zne-mir-1000-3p	2	3	1.409	ns	4	1	4.606	ns
Zne-mir-1000-5p	0	1	0.939	ns	0	0	NA	ns
Zne-mir-100-3p	0	0	NA	ns	0	0	NA	ns
Zne-mir-100-5p	0	0	NA	ns	0	0	NA	ns
Zne-mir-10-3p	0	2	0.626	ns	0	0	NA	ns
Zne-mir-10-5p	0	2	0.626	ns	0	0	NA	ns
Zne-mir-11-3p	4	0	9.394	ns	1	2	1.228	ns
Zne-mir-11-5p	0	0	NA	ns	1	1	1.842	ns
Zne-mir-1175-3p	3	2	2.505	ns	0	0	NA	ns
Zne-mir-1175-5p	0	1	0.939	ns	0	0	NA	ns
Zne-mir-12-3p	1	1	1.879	ns	0	3	0.461	ns
Zne-mir-124-3p	0	0	NA	ns	0	0	NA	ns
Zne-mir-124-5p	1	0	3.758	ns	6	2	4.299	ns
Zne-mir-125-3p	0	1	0.939	ns	0	0	NA	ns
Zne-mir-125-5p	10	0	18.788	0.0154	8	0	16.581	ns
Zne-mir-12-5p	0	0	NA	ns	0	4	0.368	ns
Zne-mir-133-3p	1	1	1.879	ns	0	1	0.921	ns
Zne-mir-133-5p	1	4	0.752	ns	3	0	7.370	ns
Zne-mir-137-3p	0	0	NA	ns	0	0	NA	ns
Zne-mir-137-5p	0	0	NA	ns	0	1	0.921	ns
Zne-mir-13a-1-3p	4	0	9.394	ns	2	0	5.527	ns
Zne-mir-13a-1-5p	0	1	0.939	ns	0	0	NA	ns
Zne-mir-13a-2-3p	4	0	9.394	ns	2	0	5.527	ns
Zne-mir-13a-2-5p	0	1	0.939	ns	0	0	NA	ns
Zne-mir-13b-3p	0	0	NA	ns	3	0	7.370	ns
Zne-mir-13b-5p	0	0	NA	ns	0	0	NA	ns
Zne-mir-1-3p	1	3	0.939	ns	0	3	0.461	ns
Zne-mir-14-3p	0	12	0.145	ns	22	0	42.375	ns
Zne-mir-14-5p	0	0	NA	ns	0	3	0.461	ns

Zne-mir-1-5p	4	0	9.394	ns	1	0	3.685	ns
Zne-mir-184-3p	0	0	NA	ns	0	0	NA	ns
Zne-mir-184-5p	1	0	1.879	7.18E-08	1	0	1.842	3.45E-05
Zne-mir-190-3p	0	0	NA	ns	0	2	0.614	ns
Zne-mir-190-5p	1	3	0.939	ns	0	0	NA	ns
Zne-mir-193-3p	1	0	3.758	ns	0	0	NA	ns
Zne-mir-193-5p	0	0	NA	ns	0	0	NA	ns
Zne-mir-210-3p	2	0	3.758	0.000942	4	0	9.212	ns
Zne-mir-210-5p	0	2	0.626	ns	1	0	3.685	ns
Zne-mir-219-3p	0	0	NA	ns	0	0	NA	ns
Zne-mir-219-5p	0	0	NA	ns	0	1	0.921	ns
Zne-mir-252-3p	0	0	NA	ns	0	1	0.921	ns
Zne-mir-252-5p	0	0	NA	ns	0	0	NA	ns
Zne-mir-263a-3p	24	0	45.092	1.11E-10	0	0	NA	ns
Zne-mir-263a-5p	2	0	5.637	ns	0	0	NA	ns
Zne-mir-263b-3p	0	1	0.939	ns	0	0	NA	ns
Zne-mir-263b-5p	0	0	NA	ns	2	1	2.764	ns
Zne-mir-275-3p	2	0	5.637	1.84E-11	9	0	16.581	0.000413
Zne-mir-275-5p	1	0	1.879	2.27E-05	0	0	NA	ns
Zne-mir-276-3p	0	0	NA	ns	0	0	NA	ns
Zne-mir-2765-3p	0	0	NA	ns	0	0	NA	ns
Zne-mir-2765-5p	0	0	NA	ns	1	0	3.685	ns
Zne-mir-276-5p	0	0	NA	ns	0	0	NA	ns
Zne-mir-277-3p	0	0	NA	ns	2	0	5.527	ns
Zne-mir-277-5p	3	0	5.637	6.23E-05	3	1	3.685	ns
Zne-mir-278-3p	0	0	NA	0.02103	1	0	3.685	ns
Zne-mir-278-5p	2	0	3.758	1.38E-09	3	0	5.527	0.01793
Zne-mir-2788-3p	0	2	0.626	ns	0	0	NA	ns
Zne-mir-2788-5p	0	0	NA	ns	0	2	0.614	ns
Zne-mir-2796-3p	5	0	9.394	0.04379	1	0	1.842	0.03191
Zne-mir-2796-5p	2	0	5.637	0.001209	0	0	NA	ns
Zne-mir-279a-3p	0	0	NA	ns	0	0	NA	ns
Zne-mir-279a-5p	1	1	1.879	1.76E-06	8	0	16.581	ns
Zne-mir-279c-3p	0	0	NA	ns	0	0	NA	ns
Zne-mir-279c-5p	4	0	7.515	5.97E-10	2	0	3.685	0.0167
Zne-mir-279d-3p	0	0	NA	ns	0	2	0.614	ns
Zne-mir-279d-5p	0	0	NA	ns	0	0	NA	ns
Zne-mir-281-3p	1	2	1.253	ns	2	1	2.764	ns
Zne-mir-281-5p	2	0	5.637	0.02979	0	1	0.921	ns
Zne-mir-282-3p	2	1	2.818	ns	0	0	NA	ns
Zne-mir-282-5p	3	2	2.818	7.37E-06	5	0	11.054	ns
Zne-mir-283-3p	0	0	NA	ns	0	0	NA	ns
Zne-mir-283-5p	0	1	0.939	ns	0	0	NA	ns

Zne-mir-29b-1-3p	0	0	NA	ns	0	0	NA	ns
Zne-mir-29b-1-5p	1	3	0.939	ns	0	2	0.614	ns
Zne-mir-2a-1-3p	0	1	0.939	ns	0	0	NA	ns
Zne-mir-2a-1-5p	2	0	5.637	ns	0	1	0.921	ns
Zne-mir-2a-2-3p	0	1	0.939	ns	0	0	NA	ns
Zne-mir-2a-2-5p	3	1	3.758	ns	1	0	3.685	ns
Zne-mir-2a-3-3p	0	1	0.939	ns	0	0	NA	ns
Zne-mir-2a-3-5p	10	0	18.788	1.39E-11	1	0	3.685	ns
Zne-mir-2a-4-3p	0	1	0.939	ns	0	0	NA	ns
Zne-mir-2a-4-5p	0	0	NA	ns	8	1	8.291	ns
Zne-mir-2b-3p	0	0	NA	ns	0	0	NA	ns
Zne-mir-2b-5p	4	0	9.394	ns	0	0	NA	ns
Zne-mir-3049-3p	2	0	5.637	0.004251	1	0	1.842	0.04929
Zne-mir-3049-5p	4	0	7.515	4.41E-06	0	0	NA	ns
Zne-mir-305-3p	0	0	NA	ns	0	0	NA	ns
Zne-mir-305-5p	0	0	NA	ns	0	0	NA	ns
Zne-mir-306-3p	0	1	0.939	ns	6	0	12.897	ns
Zne-mir-306-5p	0	0	NA	ns	0	0	NA	ns
Zne-mir-307-3p	1	0	3.758	ns	0	0	NA	ns
Zne-mir-307-5p	0	1	0.939	ns	0	2	0.614	ns
Zne-mir-31-3p	1	0	3.758	ns	2	0	5.527	ns
Zne-mir-315-3p	1	1	1.879	ns	0	0	NA	ns
Zne-mir-315-5p	0	0	NA	ns	0	0	NA	ns
Zne-mir-31-5p	1	0	3.758	ns	0	0	NA	ns
Zne-mir-316-3p	11	3	5.637	ns	0	2	0.614	ns
Zne-mir-316-5p	2	0	3.758	3.33E-05	0	0	NA	ns
Zne-mir-317-3p	2	0	3.758	ns	0	3	0.461	ns
Zne-mir-317-5p	1	0	1.879	0.01643	0	0	NA	ns
Zne-mir-33-3p	3	4	1.503	ns	5	0	11.054	ns
Zne-mir-33-5p	0	1	0.939	ns	0	0	NA	ns
Zne-mir-34-3p	55	0	103.337	8.97E-13	19	0	35.005	0.000468
Zne-mir-34-5p	0	0	NA	ns	0	0	NA	ns
Zne-mir-3477-3p	0	0	NA	ns	0	0	NA	ns
Zne-mir-3477-5p	13	14	1.754	ns	7	7	1.842	ns
Zne-mir-375-3p	0	1	0.939	ns	0	0	NA	ns
Zne-mir-375-5p	0	0	NA	ns	2	0	5.527	ns
Zne-mir-3770-3p	2	6	0.805	ns	0	0	NA	ns
Zne-mir-3770-5p	0	0	NA	ns	0	1	0.921	ns
Zne-mir-6012-3p	0	0	NA	ns	3	0	5.527	0.01583
Zne-mir-6012-5p	267	16	31.353	3.03E-15	5	9	1.105	ns
Zne-mir-71-1-3p	1	0	3.758	ns	8	0	16.581	ns
Zne-mir-71-1-5p	0	0	NA	ns	0	0	NA	ns
Zne-mir-71-2-3p	1	0	3.758	ns	8	0	16.581	ns

Zne-mir-71-2-5p	0	0	NA	ns	0	0	NA	ns
Zne-mir-7-3p	1	0	3.758	ns	0	5	0.307	ns
Zne-mir-750-3p	2	0	5.637	ns	0	0	NA	ns
Zne-mir-750-5p	1	0	3.758	ns	6	0	11.054	0.03797
Zne-mir-7-5p	0	2	0.626	ns	0	2	0.614	ns
Zne-mir-79-3p	0	0	NA	ns	2	0	5.527	ns
Zne-mir-79-5p	0	0	NA	ns	0	0	NA	ns
Zne-mir-8-3p	1	0	3.758	ns	0	0	NA	ns
Zne-mir-8-5p	1	1	1.879	ns	0	0	NA	ns
Zne-mir-87-1-3p	14	4	6.576	1.43E-05	1	2	1.228	ns
Zne-mir-87-1-5p	1	0	3.758	ns	0	0	NA	ns
Zne-mir-87-2-3p	13	3	6.576	1.43E-05	1	2	1.228	ns
Zne-mir-87-2-5p	0	0	NA	1.85E-05	0	3	0.461	ns
Zne-mir-927a-3p	0	1	0.939	0.004438	0	0	NA	ns
Zne-mir-927a-5p	0	0	NA	ns	0	1	0.921	ns
Zne-mir-927b-3p	0	0	NA	ns	0	0	NA	ns
Zne-mir-927b-5p	0	0	NA	ns	1	6	0.526	ns
Zne-mir-929-3p	0	0	NA	ns	1	0	3.685	ns
Zne-mir-929-5p	0	0	NA	ns	0	0	NA	ns
Zne-mir-92a-3p	1	0	1.879	0.01317	0	0	NA	ns
Zne-mir-92a-5p	0	0	NA	ns	0	0	NA	ns
Zne-mir-92b-3p	2	0	3.758	0.002017	0	0	NA	ns
Zne-mir-92b-5p	0	0	NA	ns	0	2	0.614	ns
Zne-mir-92c-3p	0	0	NA	ns	2	0	5.527	ns
Zne-mir-92c-5p	0	0	NA	ns	0	0	NA	ns
Zne-mir-932-1-3p	0	1	0.939	ns	0	5	0.307	ns
Zne-mir-932-1-5p	0	0	NA	ns	0	0	NA	ns
Zne-mir-932-2-3p	0	1	0.939	ns	0	5	0.307	ns
Zne-mir-932-2-5p	0	0	NA	ns	0	0	NA	ns
Zne-mir-965-3p	1	0	3.758	ns	0	0	NA	ns
Zne-mir-965-5p	0	1	0.939	ns	0	1	0.921	ns
Zne-mir-971-3p	0	0	NA	ns	0	1	0.921	ns
Zne-mir-971-5p	0	0	NA	ns	0	0	NA	ns
Zne-mir-980-1-3p	0	0	NA	ns	0	0	NA	ns
Zne-mir-980-1-5p	4	0	7.515	3.24E-05	2	0	5.527	ns
Zne-mir-980-2-3p	0	0	NA	ns	0	0	NA	ns
Zne-mir-980-2-5p	1	0	3.758	3.24E-05	2	0	5.527	ns
Zne-mir-981-3p	0	0	NA	0.0077	0	0	NA	ns
Zne-mir-981-5p	3	0	5.637	1.57E-12	1	0	3.685	1.09E-05
Zne-mir-989-1-3p	0	1	0.939	ns	0	0	NA	ns
Zne-mir-989-1-5p	0	0	NA	ns	0	0	NA	ns
Zne-mir-989-2-3p	0	1	0.939	ns	1	0	3.685	ns
Zne-mir-989-2-5p	0	0	NA	ns	0	0	NA	ns

Zne-mir-993-3p	1	0	3.758	ns	0	2	0.614	ns
Zne-mir-993-5p	2	0	5.637	ns	0	0	NA	ns
Zne-mir-995-3p	0	0	NA	ns	0	0	NA	ns
Zne-mir-995-5p	1	0	1.879	0.02062	2	0	5.527	ns
Zne-mir-998-3p	2	0	3.758	ns	0	0	NA	ns
Zne-mir-998-5p	1	0	1.879	3.56E-05	7	0	12.897	0.006444
Zne-mir-9a-3p	0	0	NA	ns	0	0	NA	ns
Zne-mir-9a-5p	0	0	NA	ns	0	0	NA	ns
Zne-mir-9c-3p	0	1	0.939	ns	0	3	0.461	ns
Zne-mir-9c-5p	0	0	NA	ns	0	0	NA	ns
Zne-mir-9d-3p	0	0	NA	ns	0	0	NA	ns
Zne-mir-9d-5p	0	1	0.939	ns	0	0	NA	ns
Zne-mir-iab-4-3p	0	0	NA	ns	1	0	3.685	ns
Zne-mir-iab-4-5p	0	0	NA	ns	2	2	1.842	ns
Zne-mir-iab-8-3p	0	0	NA	ns	1	0	3.685	ns
Zne-mir-iab-8-5p	0	0	NA	ns	1	0	3.685	ns
Totals	**556**	**136**	**7.681**		**222**	**110**	**3.718**	

Table B.17: miRNA-hit-containing DMRs are generally not 3' UTR-associated. For all DMRs that contain a significant miRNA profile hit the number falling within each genic feature is given, along with the proportion of all such DMRs this feature contains. Upstream: -1.5kb – ATG; downstream STOP - +1.5kb.

feature	count	proportion
upstream	34	0.052
exon	158	0.244
intron	401	0.619
downstream	55	0.085
Total	648	

Table B.18: gene ontology enrichment associated with miRNA-hitting DMRs relative to all DMR-containing genes

GO Term	Type	FDR	fold enrich	GO ID
regulation of small GTPase mediated signal transduction	P	3.70E-05	5.36	GO:0051056
regulation of intracellular signal transduction	P	7.70E-04	3.85	GO:1902531
response to stimulus	P	1.50E-03	1.59	GO:0050896
regulation of cell communication	P	5.10E-03	2.72	GO:0010646
regulation of signal transduction	P	5.10E-03	2.84	GO:0009966
regulation of response to stimulus	P	6.50E-03	2.55	GO:0048583
single organism signaling	P	6.70E-03	1.69	GO:0044700
GTPase regulator activity	F	6.70E-03	5.44	GO:0030695
regulation of Rho protein signal transduction	P	6.70E-03	4.53	GO:0035023
signaling	P	6.70E-03	1.67	GO:0023052
nucleoside-triphosphatase regulator activity	F	7.80E-03	4.77	GO:0060589
molecular function regulator	F	7.80E-03	2.70	GO:0098772
cell communication	P	8.10E-03	1.66	GO:0007154
signal transduction	P	9.40E-03	1.70	GO:0007165
Rho protein signal transduction	P	9.40E-03	4.21	GO:0007266
regulation of Ras protein signal transduction	P	1.20E-02	4.03	GO:0046578
cellular response to stimulus	P	1.20E-02	1.60	GO:0051716
anion binding	F	1.40E-02	1.68	GO:0043168
synapse organization	P	1.40E-02	6.64	GO:0050808
intracellular signal transduction	P	1.90E-02	2.08	GO:0035556
enzyme activator activity	F	2.00E-02	4.31	GO:0008047
anatomical structure development	P	2.00E-02	1.57	GO:0048856
guanyl-nucleotide exchange factor activity	F	2.60E-02	5.08	GO:0005085
nucleoside phosphate binding	F	3.10E-02	1.59	GO:1901265
system development	P	3.30E-02	1.63	GO:0048731
imaginal disc pattern formation	P	3.80E-02	5.37	GO:0007447
SWI/SNF superfamily-type complex	C	4.90E-02	16.11	GO:0070603
skeletal muscle fiber development	P	1.20E-01	10.25	GO:0048741
nerve development	P	1.50E-01	16.92	GO:0021675

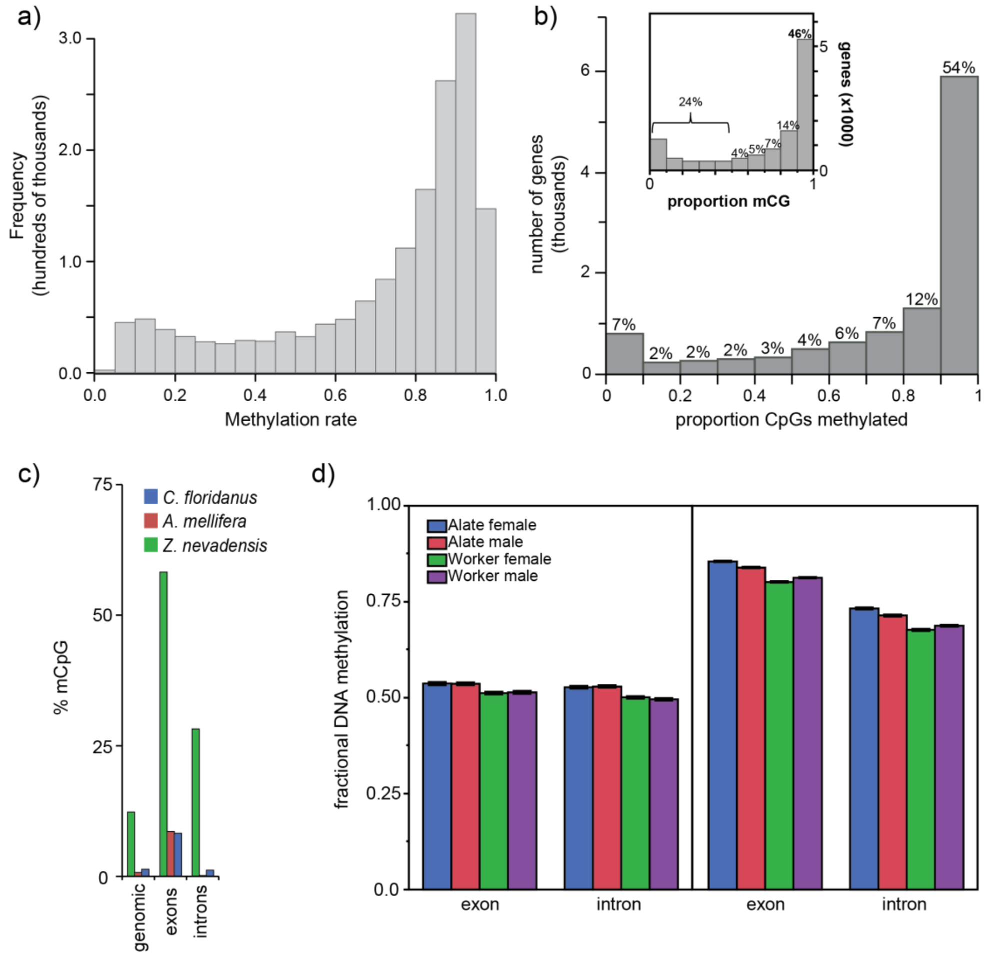


Figure B.1: **Basic DNA methylome of *Z. nevadensis*.** (a) Methylation frequency of methylated CGs in the *Z. nevadensis* genome showing that the majority of mCGs are highly methylated (>0.75). (b) Histogram of proportions of gene exonic CpGs that are mCpGs, showing >50% of methylated genes possess >=90% of CpGs methylated. (c) percentage of CpGs that are represented as mCpGs among exons, introns, and genome-wide for three species. (d) Fractional methylation level of each sample type for all exons and introns (left), as well as all exons and introns methylated in 2 or more castes/sexes (right).

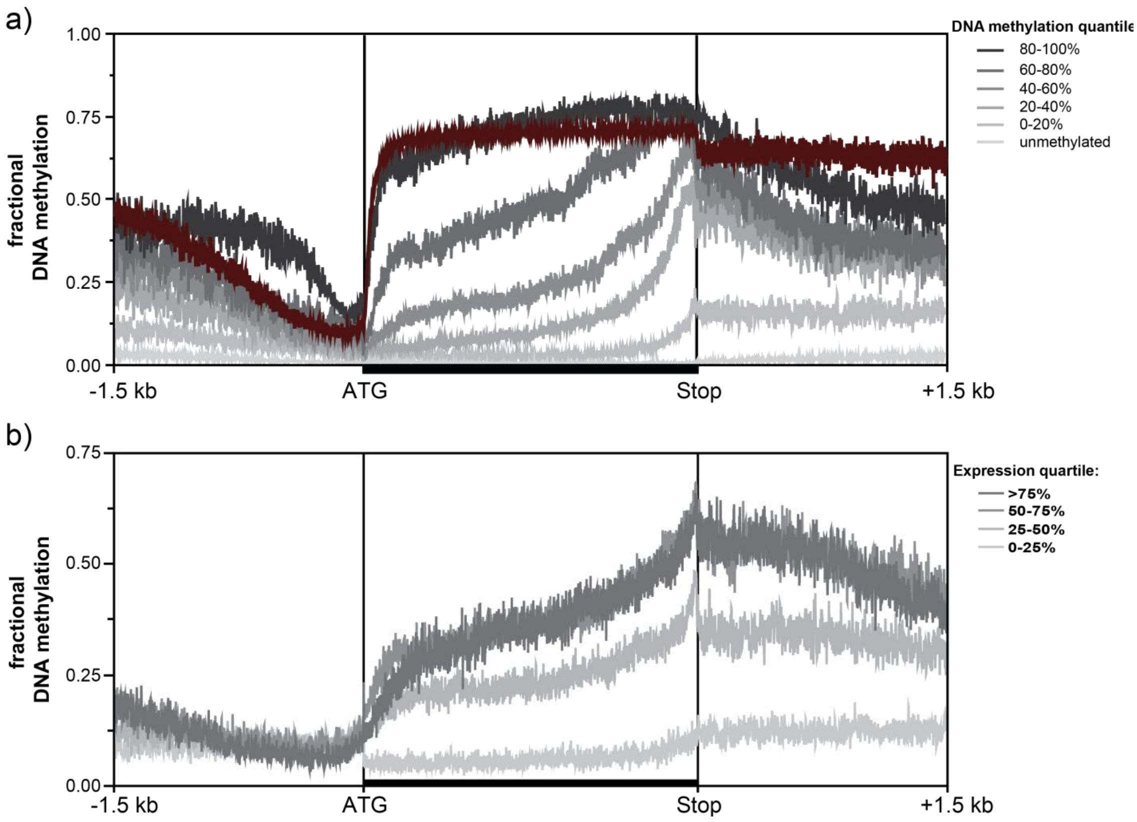


Figure B.2: **Spatial DNA methylation profiles among *Z. nevadensis* gene bodies**. (a) average positional methylation level of gene frames for quantiles of increasing methylation, as well as unmethylated genes, and human genes (maroon). (b) as in (a) but for genes of increasing expression level quantiles.

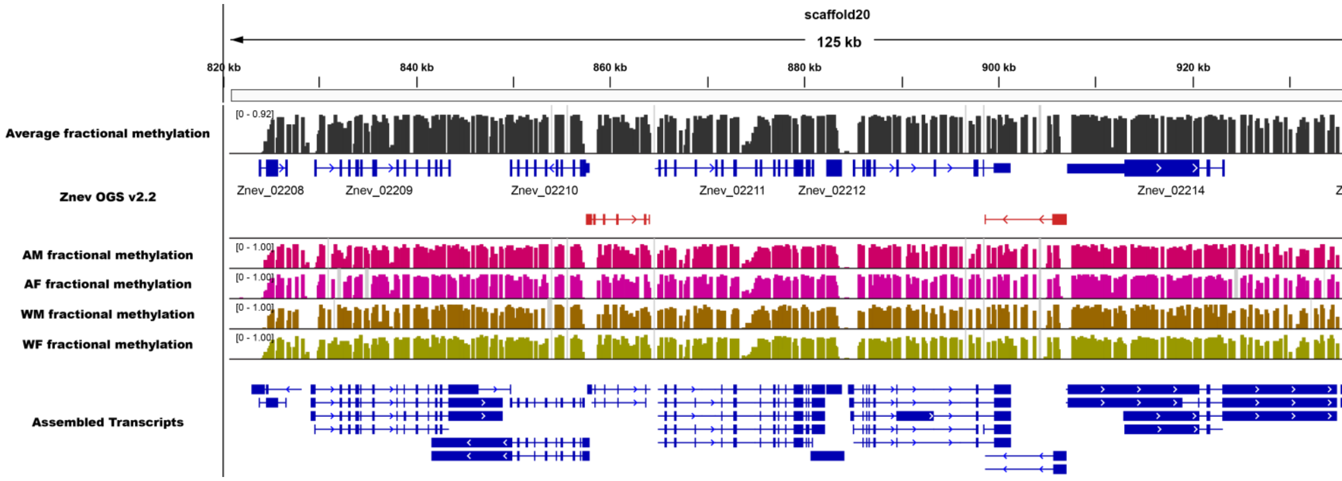


Figure B.3: **Genome browser snapshot of a >100kb high-methylation region**, showing DNA methylation as it relates to known (and novel: red) gene models across all four *Z. nevadensis* morphs.

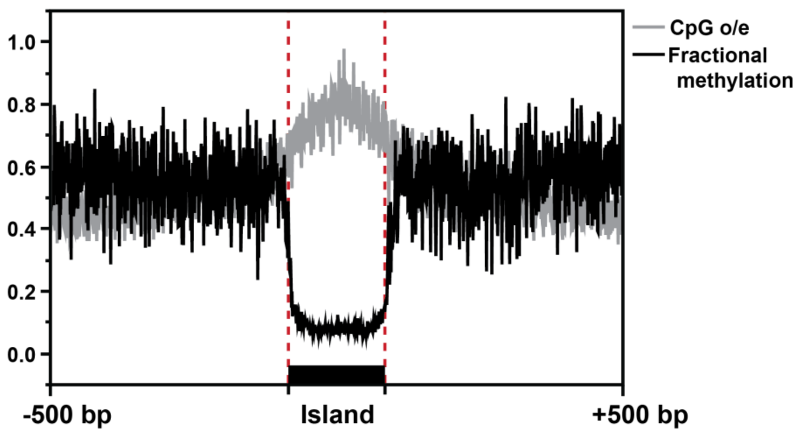


Figure B.4: **DNA methylation and CpG o/e spatial profiles over putative non-promoter (intragenic) "CpG islands"** (low-methylation regions surrounded by high methylated regions not corresponding to gene starts).

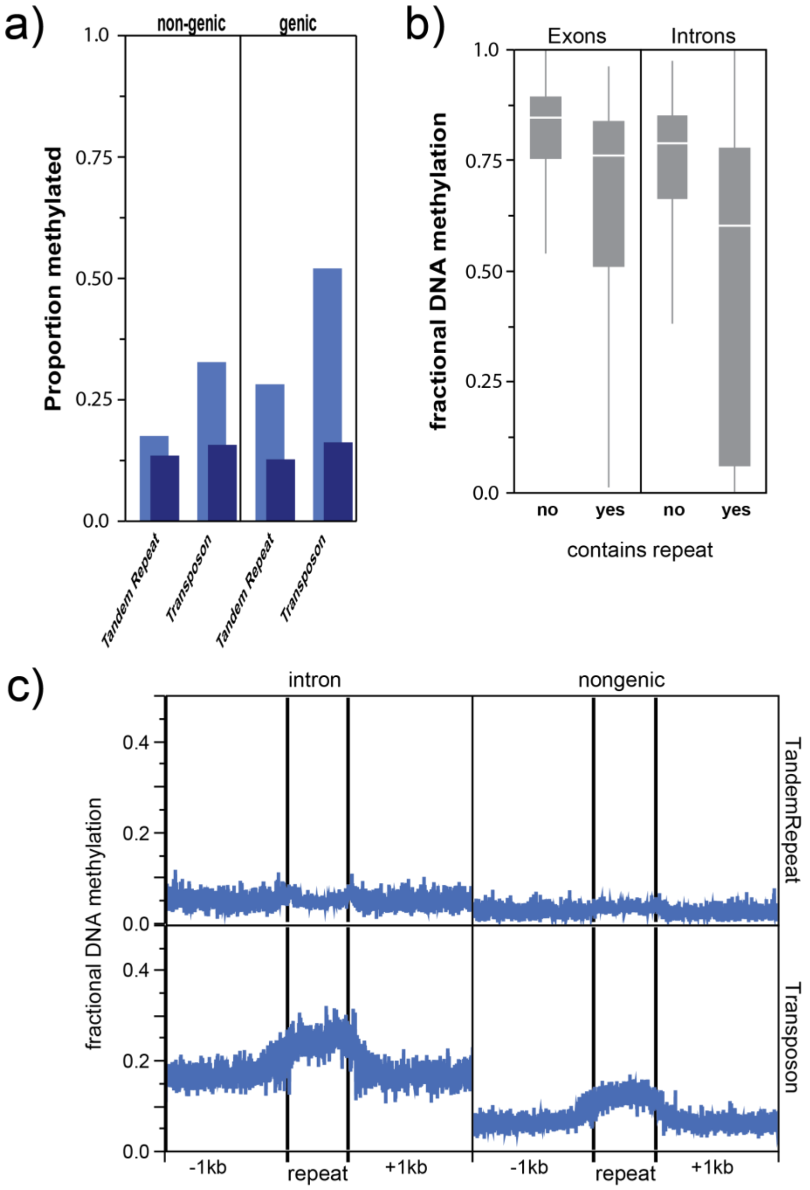


Figure B.5: **Supplementary repeat methylation analyses.** (a) proportion of each repeat type showing evidence of DNA methylation (>2 methylated CpGs) among those falling within and outside of genes, as well as for those lacking DNA methylation within the surrounding 500bp up- and down-stream (dark blue inset), (b), average methylation level of methylated exons and introns based upon whether they contain a repeat or not, and (c) spatial DNA methylation profiles within and surrounding repeats falling within introns, as well as those falling outside of genes.

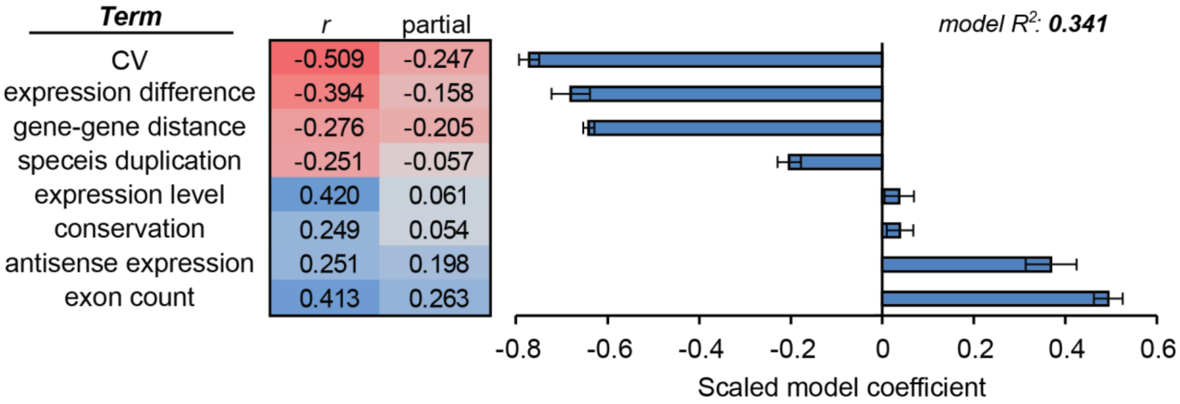


Figure B.6: **Multivariate analysis of major correlates of termite DNA methylation.** For eight variables pearson's correlation coefficients, and partial correlation coefficients are given for correlations with DNA methylation (partial coefficients: coefficients after controlling for all other variables). Also provided (bars) are scaled model coefficients from a combined regression analysis (standard least squares) of all eight variables against gene DNA methylation levels.

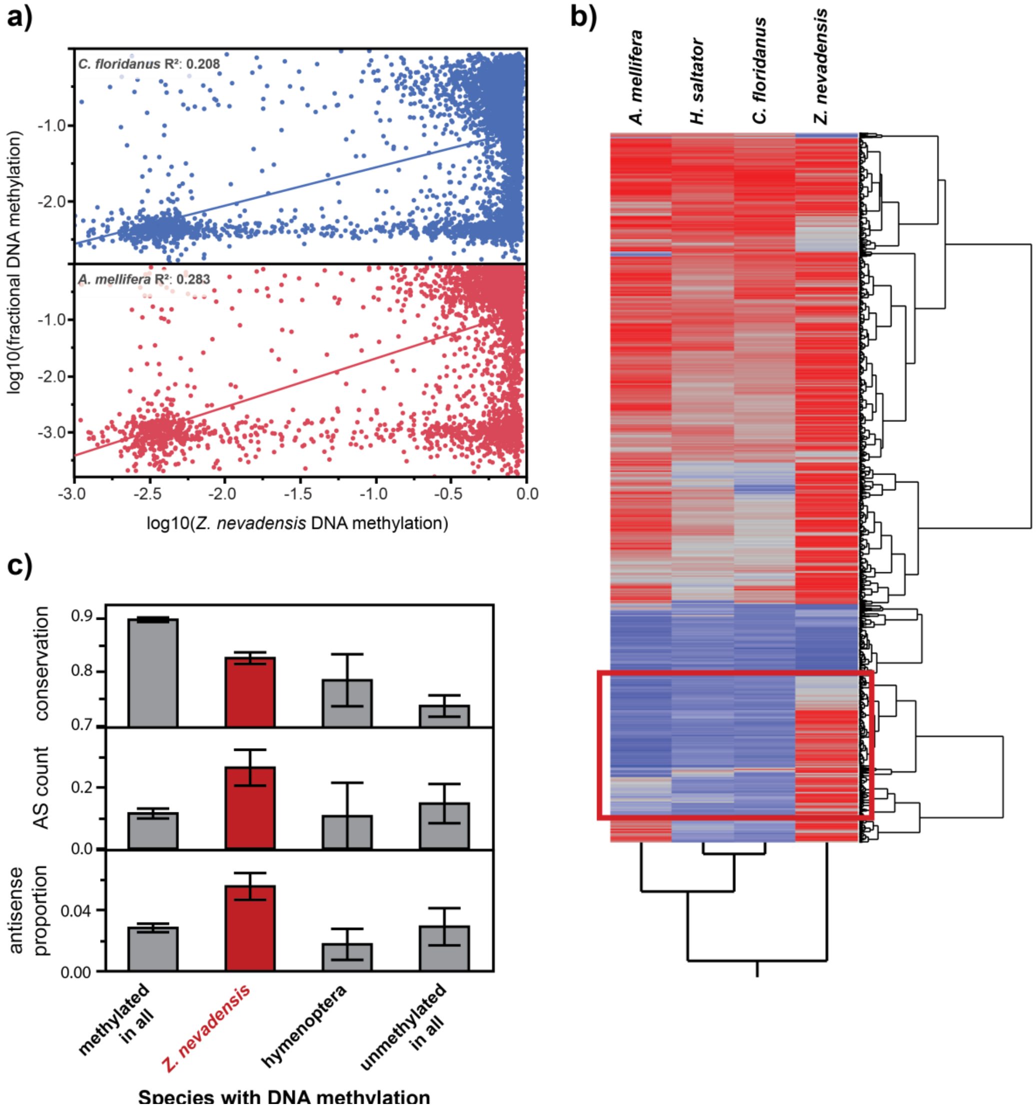


Figure B.7: DNA methylation in *Z. nevadensis* is expanded relative to hymenoptera. (a) regression between *Z. nevadensis* DNA methylation and DNA methylation levels for orthologs in *C. floridanus* (top) and *A. mellifera* (bottom) (b) Hierarchical clustering (ward method) of ~5k orthologs between *Z. nevadensis* and hymenopteran social insects illustrating large class of genes with Z.nev-specific methylation (red box), which (c) exhibit distinct qualities relative to methylated or unmethylated genes. Hymenoptera: genes methylated in ants and bees but not *Z. nevadensis*; *Z. nevadensis*+1 other: genes methylated in *Z. nevadensis* and either ants or bees (potentially indicating loss of methylation in one lineage).

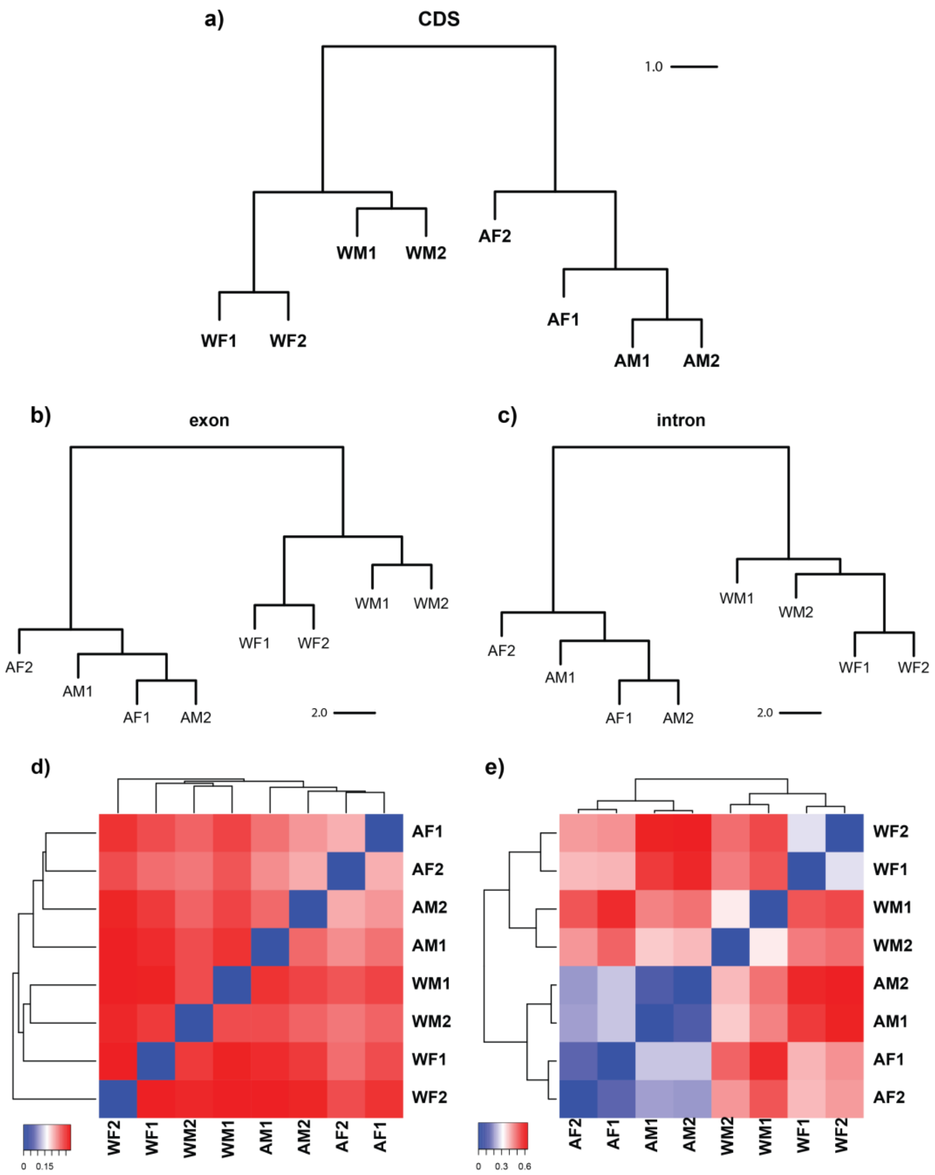


Figure B.8: **Dendrograms representing hierarchical clustering of DNA methylation libraries** based upon DNA methylation levels within a) CDS (combined exons), b) exons, c) introns, illustrating strong caste-based clustering of methylation libraries. (d) methylation library heatmaps for all mCGs shared between >4 libraries, or (e) all Differentially methylated CpGs (DMCs).

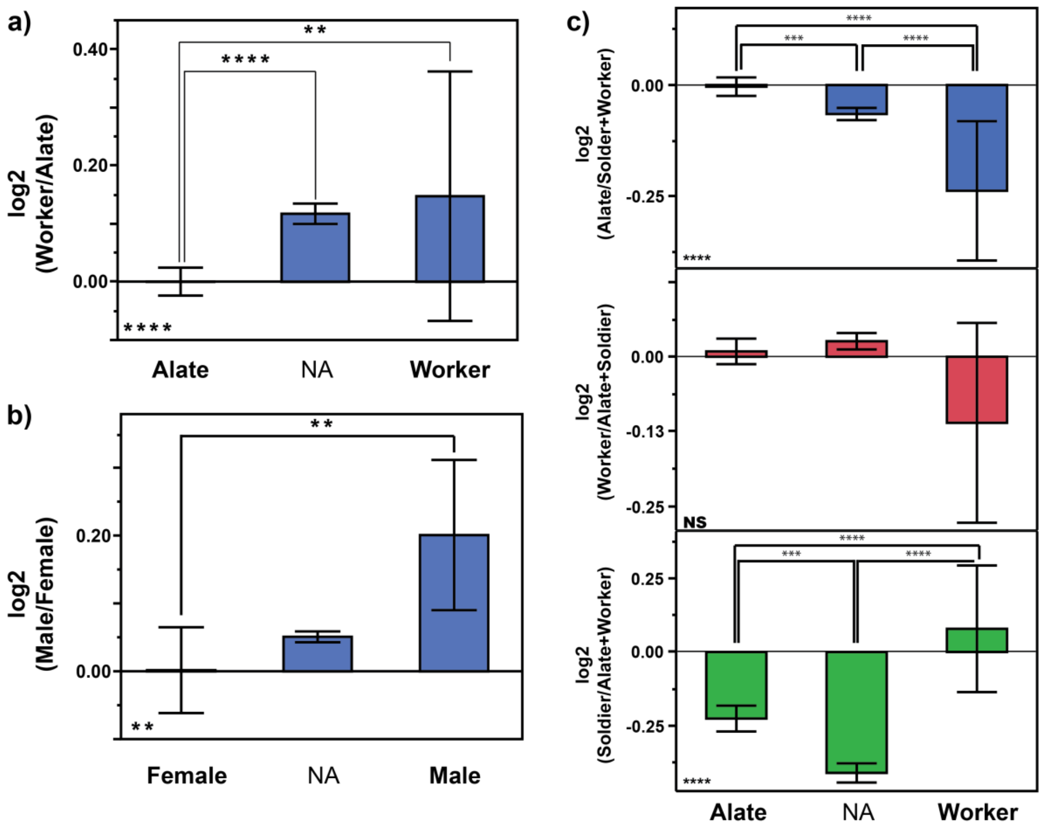


Figure B.9: **Differentially metylated genes show weak signal of caste-of-hypermethylation increased expression.** Ratio of gene expression bias by DMG up-methylation type for (a) worker/alate expression ratio by caste biased DMG type, and (b) male/female expression ratio by sex biased DMG type. (c) Caste biased DMGs were also compared to ratios of gene expression single caste-specific bias for each of three castes: Alate (top), Worker (middle), and Soldier (bottom). Bottom left of each graph features Kruskal-Wallis significance test Pvalue. All other pavlues related to wilcoxon *post hoc* pairwise tests (assuming full test significance).

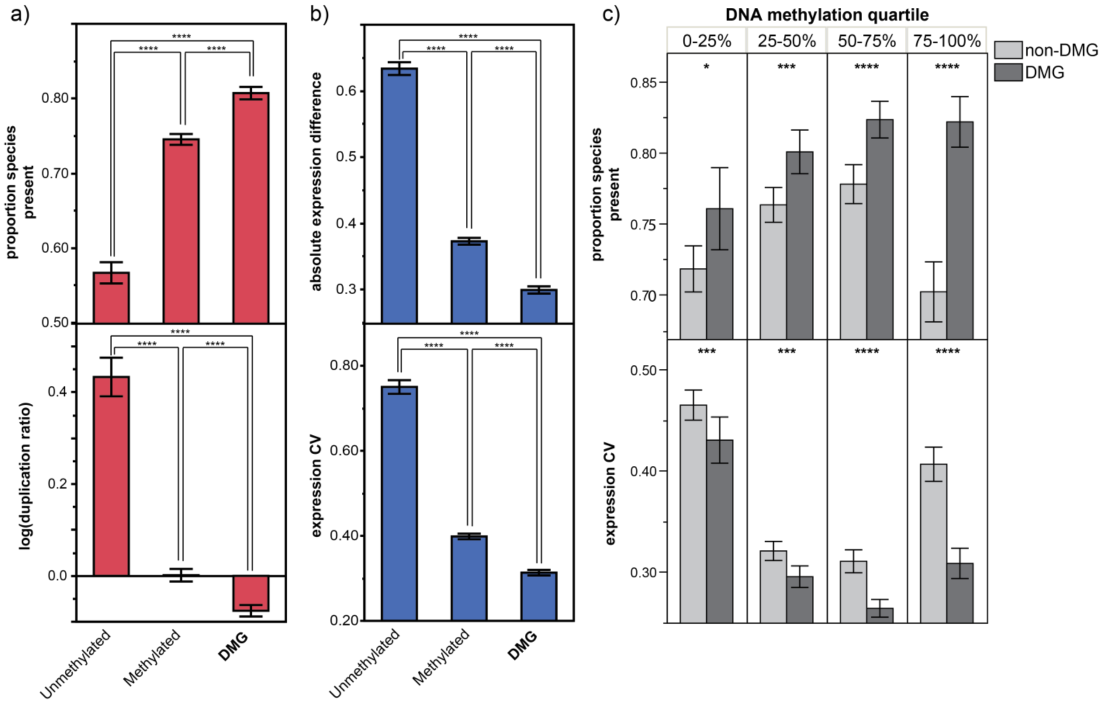


Figure B.10: Differentially methylated genes show distinct expression and evolutionary conservation. (a) Gene conservation (top) and *Z. nevadensis* duplication ratio (*Z. nevadensis* orthodb copy number/insect-wide orthodb copy number; bottom), as well as (b) absolute gene expression difference between 4 morphs (top) and gene expression noise (bottom) is presented for unmethylated, methylated (but not differing), and differentially methylated genes. (c) Expression CV and gene conservation presented for differentially methylated genes (DMG) and non-differentially methylated genes (non-DMG) across four quartiles of DNA methylation level, showing DMGs differ from non-DMGs consistently across methylation levels. Pvalues from (a) and (b) represent results from Wilcoxon *post hoc* pairwise tests (all comparisons significant at group level), and (c) Wilcoxon rank-sum tests.

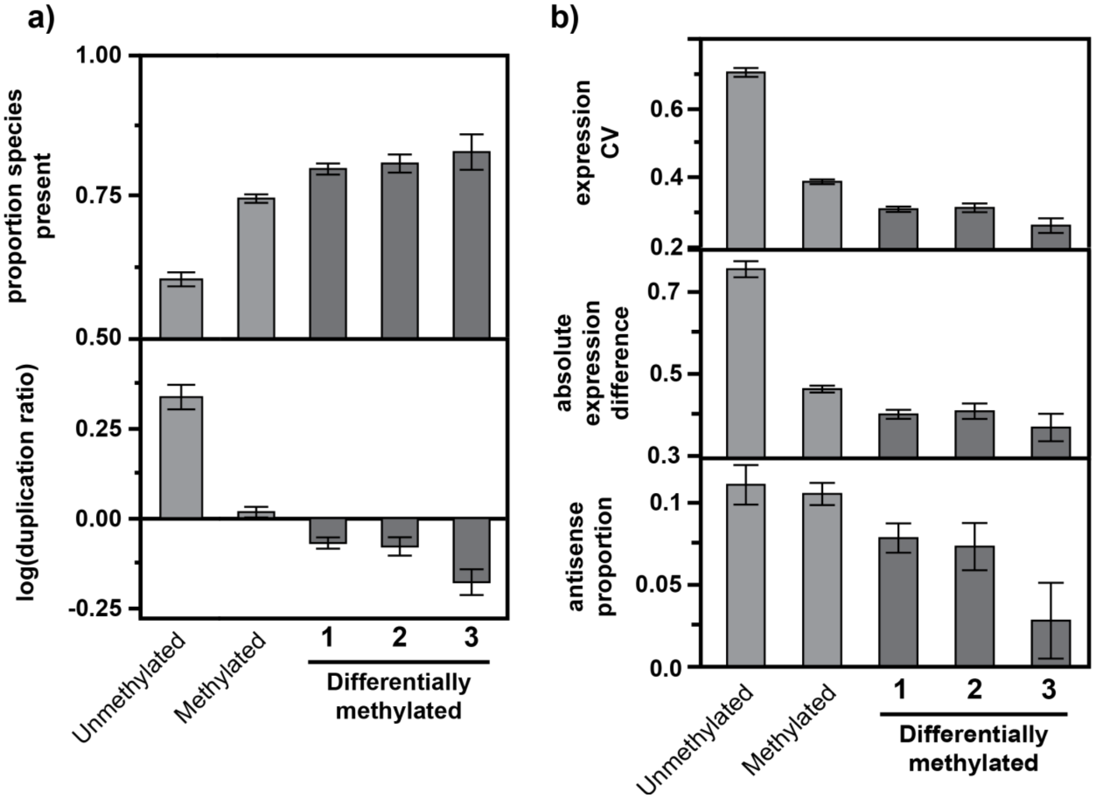


Figure B.11: Differentially methylated genes are more conserved and less variably expressed than unmethylated or (non-differentially) methylated genes. a) Proportion of species with a representative orthodb ortholog group member present (top) as well as the ratio of *Z. nevadensis* copy number to average copy number across insect species (bottom) is given for differentially methylated genes exhibiting differential methylation in one, two, or three or more (1-3 respectively) pairwise tests (of 4), as well as non-differing methylated genes, and unmethylated genes. b) gene expression coefficient of variation (top), absolute sample expression fold change (middle), and proportion of gene reads that map to antisense strand (bottom) for the same as in a).

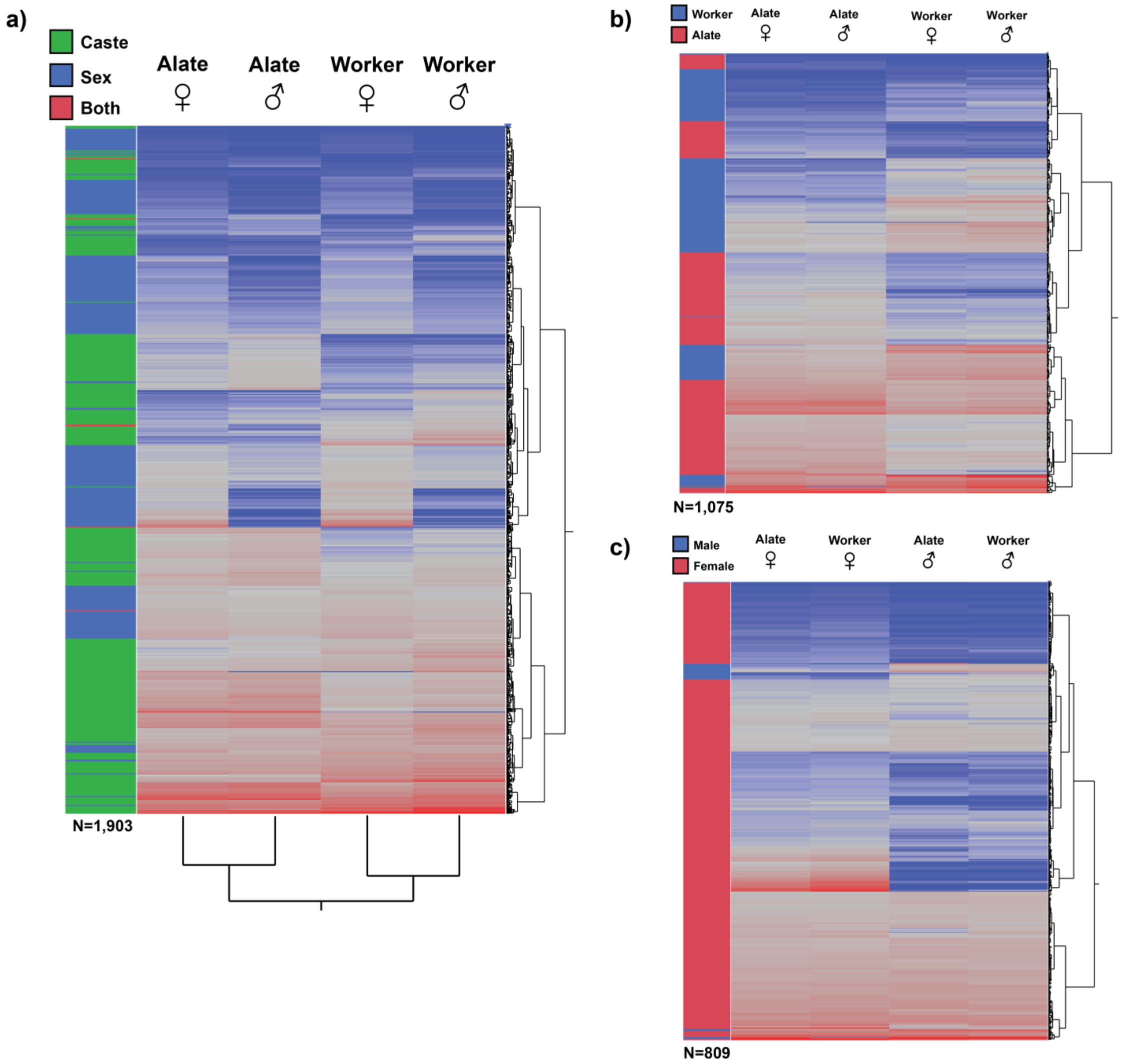


Figure B.12: hierarchical clustering of genes showing evidence of differential expression between castes or sexes. a) all genes showing differential expression between either castes or sexes. b) Clustering of only genes differentially expressed between castes, and c) sexes.

APPENDIX C

SUPPLEMENTARY MATERIAL FOR CHAPTER 4

Supplementary tables and figures:

Table C.1: **Percentages of genes that were significantly methylated and also marked by a given hPTM, as well as over- or under-representation of hPTM enrichment among methylated genes** as determined by a Fisher's exact test (all Fisher's exact test P values << 0.0001).

	% methylated that are also marked by hPTM	% marked by hPTM that are also methylated	Fisher's exact test significant direction
H3	81.1	70.8	over
H3K4me3	80.6	79	over
H3K4me1	27.2	80	over
H3K27me3	1.2	23.4	under
H3K27ac	85.8	69.5	over
H3K36me3	79.6	79.6	over
H3K9me3	1.5	3.3	under
H3K9ac	67.3	42.1	over
PolII	49.5	66.1	over

Table C.2: **Spearman's rank correlations between fractional DNA methylation and histone modification normalized tag enrichment at genic features** ("TSS-proximal" represents a 2kb length-normalized gene measure 500bp upstream of TSS to 1.5kb downstream of gene start). P < 0.0001 for all listed correlations.

	TSS-proximal	**Exon**	**Intron**
H3	0.599	0.307	0.185
H3K4me3	0.621	0.417	0.290
H3K4me1	0.158	0.361	0.182
H3K27me3	0.162	-0.086	-0.143
H3K27ac	0.596	0.376	0.349
H3K36me3	0.617	0.459	0.202
H3K9me3	-0.564	-0.448	-0.319
H3K9ac	0.272	-0.153	-0.052
PolII	0.464	0.121	0.134

Table C.3: **Numbers of HMRs associated with specific gene features**. Genic: intersecting any gene annotation (gene set model or valid cufflinks transcript). Proximal: falling within 2kb either up- (5') or downstream (3') of any gene annotation. Non-genic: HMRs not falling within 2kb of a gene annotation. Non-genic HMRs were further divided into those which showed experimental evidence of expression (>4 RNA-sequencing reads mapped to HMR) despite the lack of a gene annotation, and those without (without RNA-seq).

Feature	**HMR count**
Genic	6927
exonic	*4447*
Intronic	*2480*
Proximal	433
5'	*100*
3'	*333*
Non-genic	22
with RNA-seq	*8*
without RNA-seq	*14*
Total	**7382**

Table C.4. **Association tests between a genomic region's differential methylation status (whether it is a DMR or unchanging methylated region) and differential ChIP enrichment (differentially enriched between castes or not) for the 8 factors assessed in this study among windows with sufficient DNA methylation and differential enrichment data.**

	DMR status	**% nonDiffChip**	**% DiffChip**	**N**	**P value**
H3K27ac	nonDMR	48.44	51.56	5850	NS
	DMR	47.98	52.02	2728	
H3K27me3	nonDMR	60	40	597	NS
	DMR	55.88	44.12	300	
H3K36me3	nonDMR	29.65	70.35	5501	<0.0001
	DMR	25.61	74.39	2722	
H3K4me1	nonDMR	85.9	14.1	2128	NS
	DMR	86.22	13.78	849	
H3K4me3	nonDMR	43.9	56.1	5917	<0.0001
	DMR	38.63	61.37	3219	
H3K9ac	nonDMR	69.69	30.38	2946	0.0031
	DMR	65.21	34.79	1602	
H3K9me3	nonDMR	58.11	41.89	101	NS
	DMR	55.56	44.44	85	
PolII	nonDMR	47.31	52.69	4170	NS
	DMR	48.15	51.85	1942	

NS, non-significant

Table C.5: **Comparisons between differentially methylated regions and overlapping differentially enriched ChIP calls** (using standard data from ‘consolidatedCls’ file). P value from likelihood ratio test. Percentages reflect the percent membership a given row shows in the associated column.

			Differential methylation call			
			Male	**Worker**	**NA**	**P value**
Differential ChIP enrichment call	**H3K27ac**	% Male	46.23	54.3	50.78	0.0004
		% Worker	53.77	45.7	48.22	
		N	928	1606	4525	
	H3K27me3	% Male	95.52	92.47	92.33	NS
		% Worker	4.48	7.53	7.67	
		N	67	93	313	
	H3K36me3	% Male	88.25	89.5	87.4	0.0068
		% Worker	11.75	10.5	12.6	
		N	2196	3112	7981	
	H3K4me1	% Male	30.77	32.65	31.21	NS
		% Worker	69.23	67.35	68.79	
		N	91	147	487	
	H3K4me3	% Male	78.45	84.97	78.71	<0.0001
		% Worker	21.55	15.03	21.29	
		N	815	1942	5077	
	H3K9ac	% Male	90.97	80.85	83.1	0.0145
		% Worker	9.03	19.15	16.9	
		N	144	329	728	
	H3K9me3	% Male	40.48	26.83	33.33	NS
		% Worker	59.52	73.17	66.67	
		N	42	41	87	
	PolII	% Male	54.5	63.2	59.1	0.0010
		% Worker	45.5	36.8	40.9	
		N	556	1333	3579	

NS, non-significant

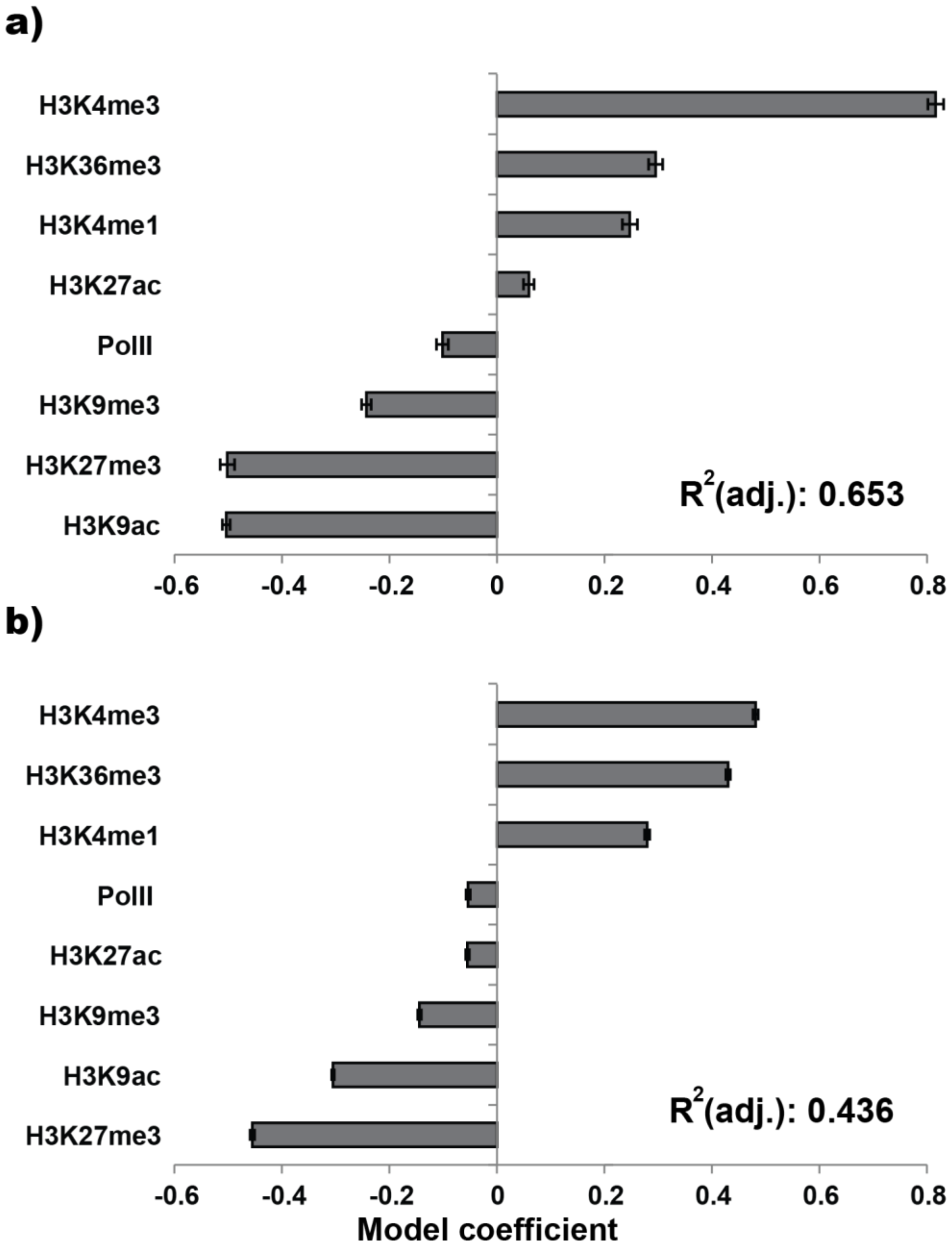


Figure C.1: **hPTM levels explain DNA methylation variation.** Model coefficients for multiple regression of hPTM enrichment levels against DNA methylation levels within the same feature for a) CDS, and b) exons+introns (as distinct features) as the dependent variable. Magnitude of bars represent estimated model coefficients. Interaction terms not included. Error bars represent standard error. R^2 values given represent adjusted R^2 for full model fit.

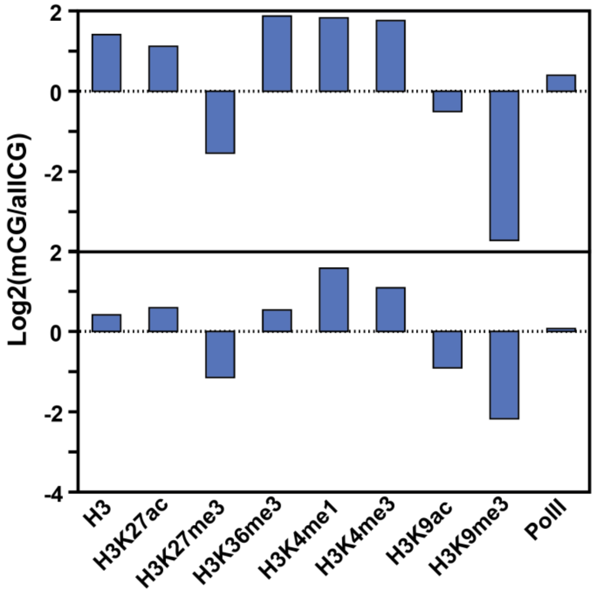


Figure C.2: Methylated CpGs are strongly over- or under-represented among regions significantly enriched for different hPTMs. For both a) within all gene bodies, as well as b) within only methylated gene bodies the log2-transformed ratio of the proportion of methylated CpGs (mCGs) falling within the given hPTM-enriched regions to that of of all CpGs (allCGs) falling within the same regions is given.

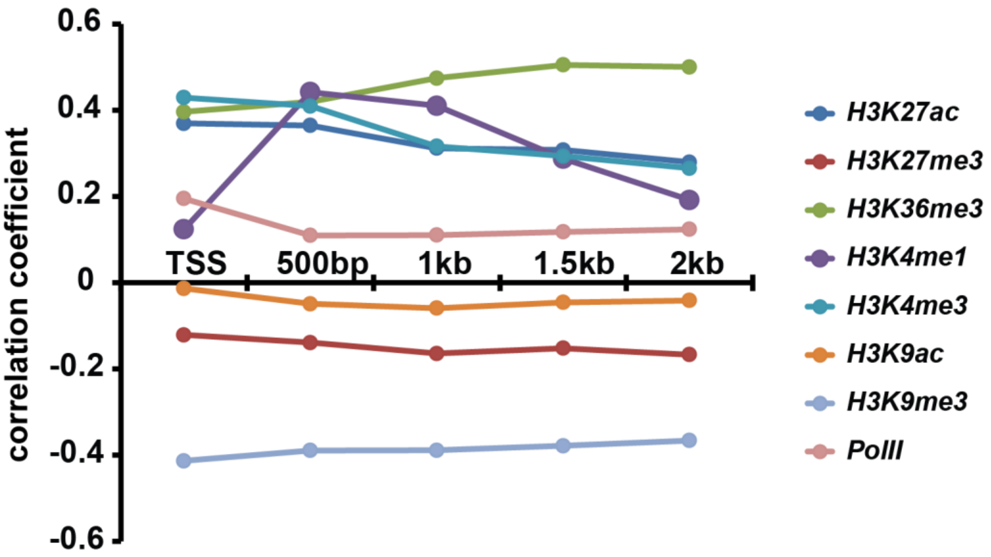


Figure C.3. DNA methylation is correlated with hPTM enrichment at a fine spatial scale within genes. The correlation coefficients for spearman's rank correlations between DNA methylation and hPTM enrichment for each hPTM for 500bp windows downstream of gene TSSs are shown. Each point represents the correlation between DNA fractional methylation and the given hPTM tag fold enrichment within a 500bp window starting the given distance (x axis) from the TSS (eg TSS=0-500bp from start of TSS). Only genes longer than 2.5kb were used. All correlations are significant (P<0.05).

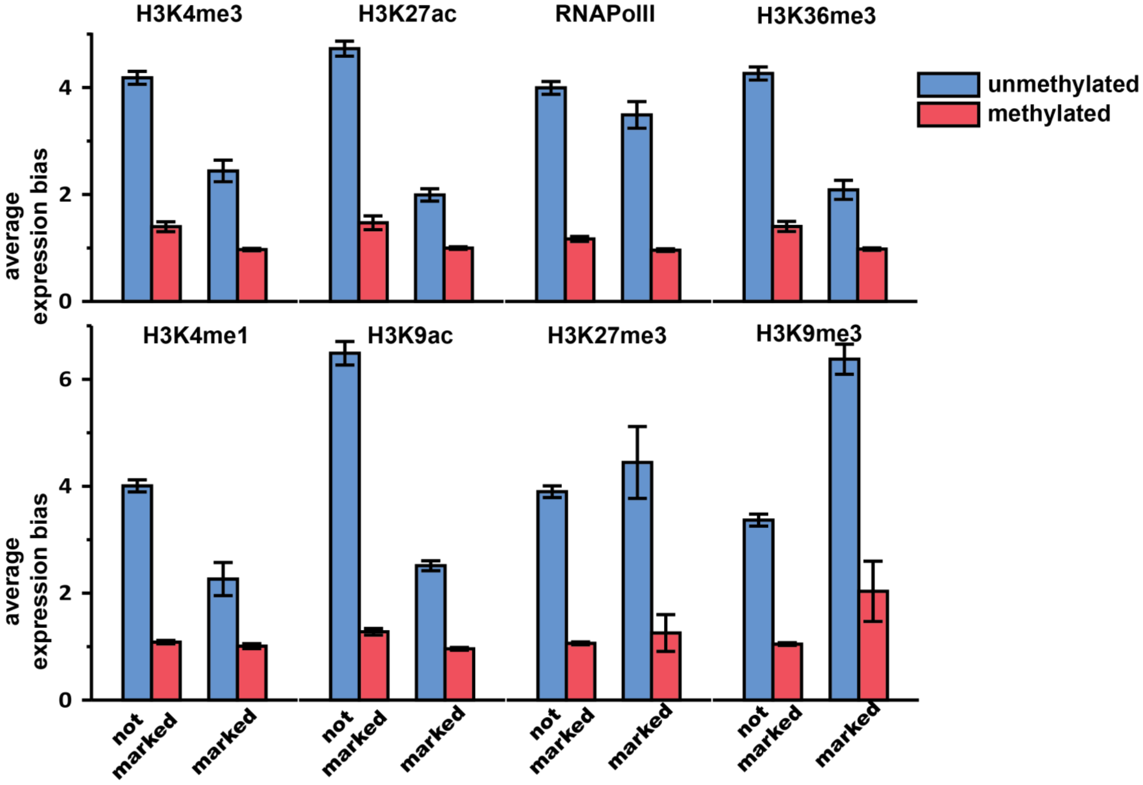


Figure C.4: **Levels of expression bias (average of absolute $\log_2$(FPKM) ratios for 3 comparisons) of genes associated with histone modifications.** Methylated genes exhibit consistently lower levels of expression bias relative to unmethylated genes.

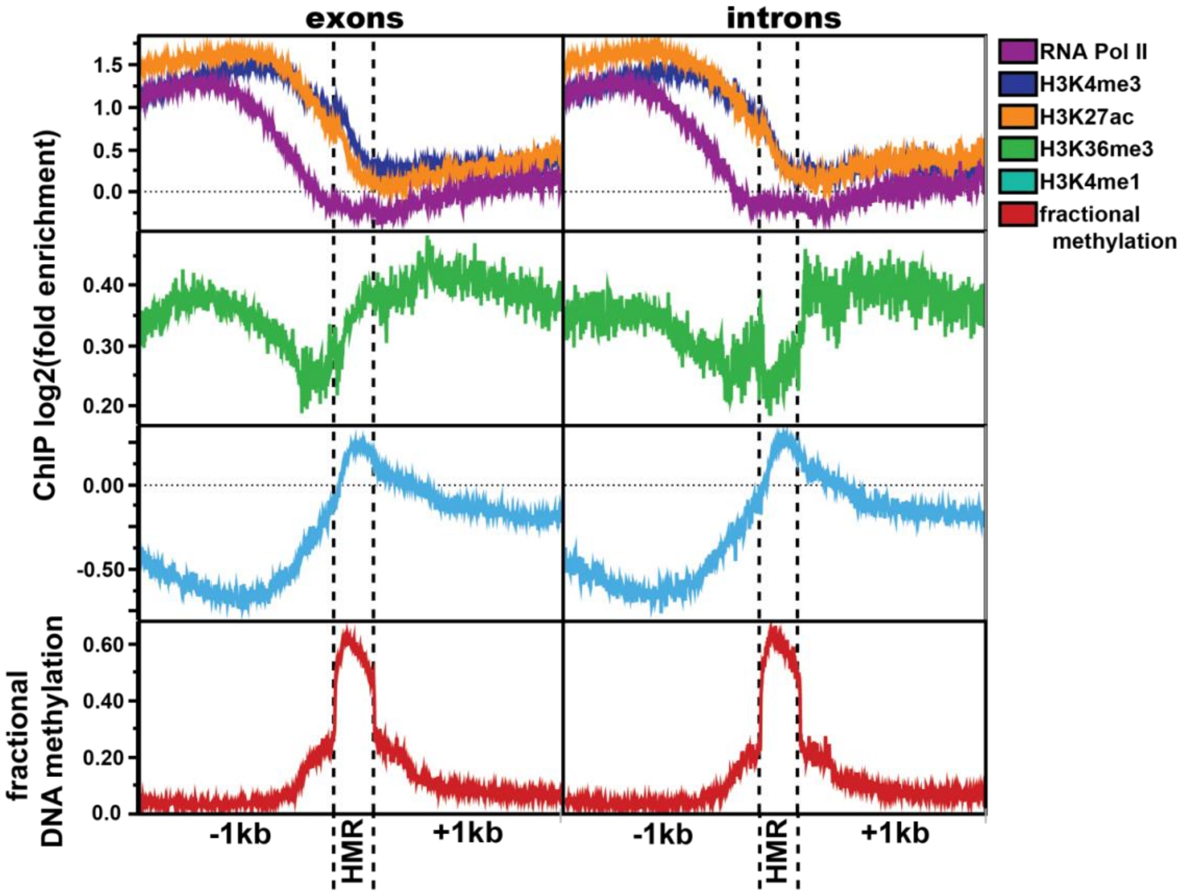


Figure C.5: **Chip profiles as they relate to highly methylated regions (HMRs) localized to exons and introns.** Shown ChIP measures correspond to those in Figure 3 of the main text.

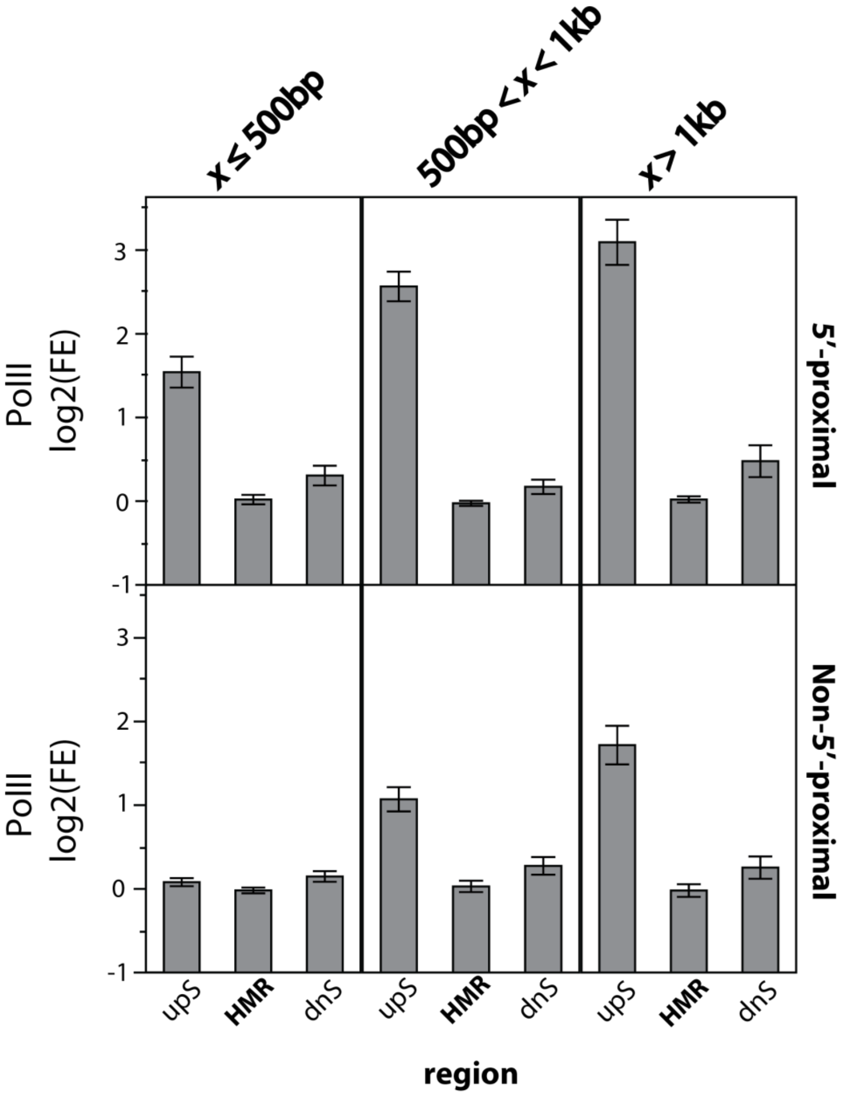


Figure C.6: **RNA polymerase II ChIP enrichment (log2 fold enrichment over input) within highly methylated regions (HMRs)** as well as 1kb regions in the 5' and 3' directions (upS and dnS, respectively), split both by HMR proximity to a gene start, as well as grouped into 3 HMR length classes. All comparisons are significant below the p<0.0001 level.

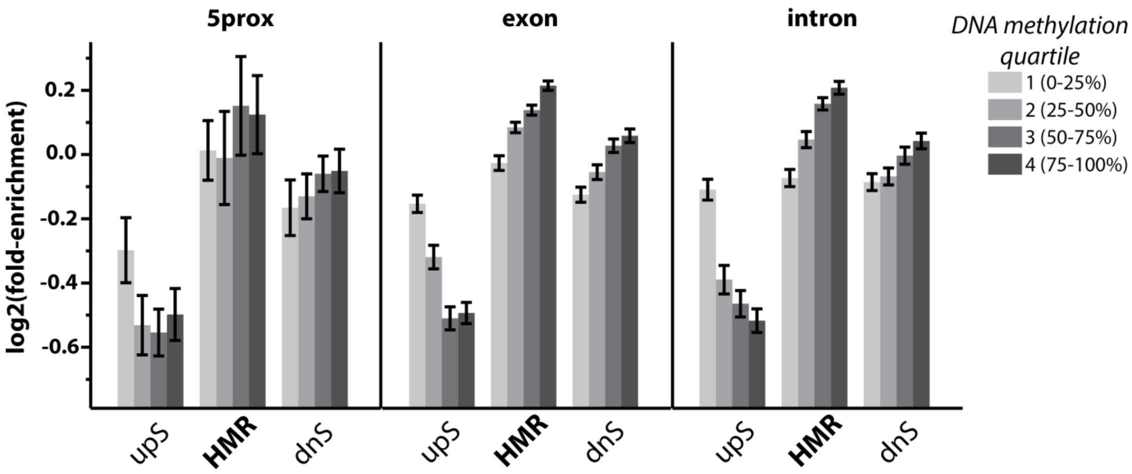


Figure C.7: **H3K4me1 is enriched within highly methylated regions (HMRs), independent of genic context.** H3K4me1 values shown for HMRs, and 1kb bins up- and downstream (upS and dnS, respectively) of HMR for HMRs overlapping the region -2kb-0kb from gene starts (5prox), exons, and introns. HMR DNA methylation levels were split into 4 equally-sized ascending quartiles to illustrate opposite relationship between DNA methylation level and H3K4me1 enrichment between regions upstream of HMRs and HMRs themselves.

APPENDIX D
SUPPLEMENTARY MATERIAL FOR CHAPTER 5

Supplementary Text

Common factors associated with coding sequence evolution in two insect orders

Our investigation into genetic and epigenetic factors associated with rates of coding sequence evolution provided insight into the factors that shape evolutionary rate variation in two insect orders, the Diptera and the Hymenoptera, that diverged approximately 350 Ma (Wiegmann, Trautwein et al. 2009). In order to directly assess common factors associated with variation in the evolutionary rates of coding sequences, we generated multiple linear regression models using only those characteristics of orthologs for which data were present in both *C. floridanus* and *D. melanogaster* (Table D.3). We found that average exon length was positively associated with both nonsynonymous substitution rate (dN) and synonymous substitution rate (dS) in *C. floridanus* and *D. melanogaster* (Table D.3). This previously discovered relationship may be explained by the associations of mean exon size with gene expression breadth (Duret and Mouchiroud 2000; Eisenberg and Levanon 2003) and nucleosome positioning (Schwartz, Meshorer et al. 2009; Prendergast and Semple 2011; Lawrie, Messer et al. 2013). Similarly, the number of introns in a gene was positively associated with dN when considered in a multiple regression framework (Table D.3), consistent with selection for compactness operating on highly conserved genes (Eisenberg and Levanon 2003). In contrast, intron length was negatively associated with dN/dS (Table D.3), thus highlighting complexities in the relationship between gene compactness and selection (Carmel, Rogozin et al. 2007; Carmel and Koonin 2009).

We found that gene expression level was negatively associated with dN and dN/dS in both *C. floridanus* and *D. melanogaster* (Table D.3). This relationship has been widely observed and may be attributable to selection against protein mistranslation

(Duret and Mouchiroud 2000; Drummond, Raval et al. 2006; Drummond and Wilke 2008). Similarly, two histone modifications associated with active transcription, H3K4me3 and H3K36me3 (Kharchenko, Alekseyenko et al. 2011; Zhou, Goren et al. 2011), were negatively associated with dN/dS. Moreover, H3K4me3 was positively associated with dS in both taxa (Table D.3), suggesting this epigenetic mark may be linked to variation in mutation rate or structural constraints on chromatin (Prendergast, Campbell et al. 2007; Park, Qian et al. 2012).

Supplementary Tables and Figures

Table D.1. Pearson's correlations between *C. floridanus* gene characteristics and dS, as compared to correlations with dS after masking CpG sites

X	*X* correlation with ant dS	*X* correlation with ant dS, CpGs masked	Percent decrease in correlation after CpG masking	P-value, dS correlation	P-value, dS, CpGs masked correlation
DNA methylation	0.28	0.15	46%	< 0.0001	< 0.0001
Exon length (mean)	0.28	0.20	27%	< 0.0001	< 0.0001
Expression level	0.08	0.05	31%	< 0.0001	0.0008
H3K27ac	0.15	0.10	34%	< 0.0001	< 0.0001
H3K27me3	-0.09	-0.03	68%	< 0.0001	0.0660
H3K36me3	0.05	0.00	93%	0.0020	0.8225
H3K4me1	0.05	0.01	81%	0.0047	0.5926
H3K4me3	0.19	0.14	29%	< 0.0001	< 0.0001
H3K9ac	-0.21	-0.11	48%	< 0.0001	< 0.0001
H3K9me3	-0.17	-0.07	57%	< 0.0001	< 0.0001
Intron length (mean)	-0.25	-0.15	39%	< 0.0001	< 0.0001
Intron count	-0.26	-0.19	28%	< 0.0001	< 0.0001
RNA Pol II	0.17	0.13	25%	< 0.0001	< 0.0001

Table D.2. Gene ontology annotation enrichment of genes with the highest 300 values for each *Camponotus floridanus* principal component (from the analysis summarized in Table 5.1) relative to all other genes in our principal component analysis.

PC	Term	Category[a]	Accession	No. genes (of 300)	Fold-enriched	FDR *P*-value
PC 1	structural constituent of ribosome	F	GO:0003735	17	6.0	7.74E-04
	ribosome biogenesis	P	GO:0042254	19	4.3	2.43E-03
	proteasome complex	C	GO:0000502	10	8.1	2.82E-03
	translation	P	GO:0006412	27	2.8	5.28E-03
	actin polymerization or depolymerization	P	GO:0008154	7	12.4	6.13E-03
	proton transport	P	GO:0015992	10	6.7	6.13E-03
	oxidative phosphorylation	P	GO:0006119	9	6.6	1.16E-02
	cytosolic ribosome	C	GO:0022626	3	NA[b]	3.72E-02
	respiratory chain complex	C	GO:0098803	6	9.5	3.72E-02
	hydrogen ion transmembrane transporter activity	F	GO:0015078	8	5.8	4.79E-02
PC 2	none significant at FDR *P* < 0.05					
PC 3	sequence-specific DNA binding	F	GO:0043565	19	3.8	1.72E-02
	aromatic compound biosynthetic process	P	GO:0019438	48	2.2	1.72E-02
	cellular nitrogen compound biosynthetic process	P	GO:0044271	48	2.2	1.72E-02
	heterocycle	P	GO:00181	48	2.1	1.72E-

biosynthetic process		30			02
organic cyclic compound biosynthetic process	P	GO:1901362	49	2.1	1.72E-02
regulation of transcription, DNA-templated	P	GO:0006355	26	2.8	1.72E-02
sequence-specific DNA binding transcription factor activity	F	GO:0003700	21	3.0	2.66E-02
regulation of transcription, DNA-templated	P	GO:0006355	32	2.3	2.99E-02

[a] P, biological process; F, molecular function; C, cellular component

[b] No genes with this term were present in the reference set, preventing calculation of fold enrichment.

Table D.3. Similarities and differences in associations with coding sequence evolution in an ant (*C. floridanus*) and a fly (*D. melanogaster*) in linear models generated from traits with data in both taxa, limited to common orthologs (n = 2102)

	***C. floridanus* model coefficient**	***D. melanogaster model coefficient***	**Consistent significant relationship between species?**
X	**dS (R^2 = 0.25)**	**dS (R^2 = 0.35)**	
H3K9ac	-0.41****	-0.05	No
Intron count	-0.33****	0.59****	No
H3K4me1	-0.05*	-0.14****	Yes, negative
H3K36me3	-0.05	0.10**	No
H3K27me3	-0.03	-0.03	No
Expression level	0.01	-0.27****	No
H3K9me3	0.02	0.11****	No
H3K27ac	0.04	0.01	No
RNA Pol II	0.09**	-0.07*	No
Intron length	0.21***	-0.54****	No
H3K4me3	0.27****	0.52****	Yes, positive
Exon length	0.28****	0.37****	Yes, positive
X	**dN (R^2 = 0.16)**	**dN (R^2 = 0.16)**	
H3K9ac	-0.24****	0.07	No
H3K36me3	-0.17****	-0.07	No
Intron length	-0.09	-0.42****	No
Expression level	-0.06*	-0.23****	Yes, negative
H3K4me3	-0.04	0.02	No
H3K9me3	-0.02	0.03	No
H3K4me1	0.05	0.00	No
H3K27ac	0.12**	0.01	No
RNA Pol II	0.14****	-0.04	No
Intron count	0.16*	0.54****	Yes, positive
H3K27me3	0.17****	-0.14****	No
Exon length	0.38****	0.36****	Yes, positive
X	**dN/dS (R^2 = 0.12)**	**dN/dS (R^2 = 0.08)**	
Intron length	-0.17**	-0.26**	Yes, negative
H3K36me3	-0.16****	-0.11**	Yes, negative
H3K4me3	-0.14***	-0.17**	Yes, negative
H3K9ac	-0.11***	0.09	No
Expression level	-0.07*	-0.15****	Yes, negative
H3K9me3	-0.03	0	No
H3K4me1	0.07**	0.04	No
H3K27ac	0.11**	0.01	No
PolII	0.11***	-0.01	No
H3K27me3	0.19****	-0.14****	No

Intron count	0.30****	0.37****	Yes, positive
Exon length	0.31****	0.25****	Yes, positive

*P < 0.05, **P < 10-2, ***P < 10-3, ****P < 10-4

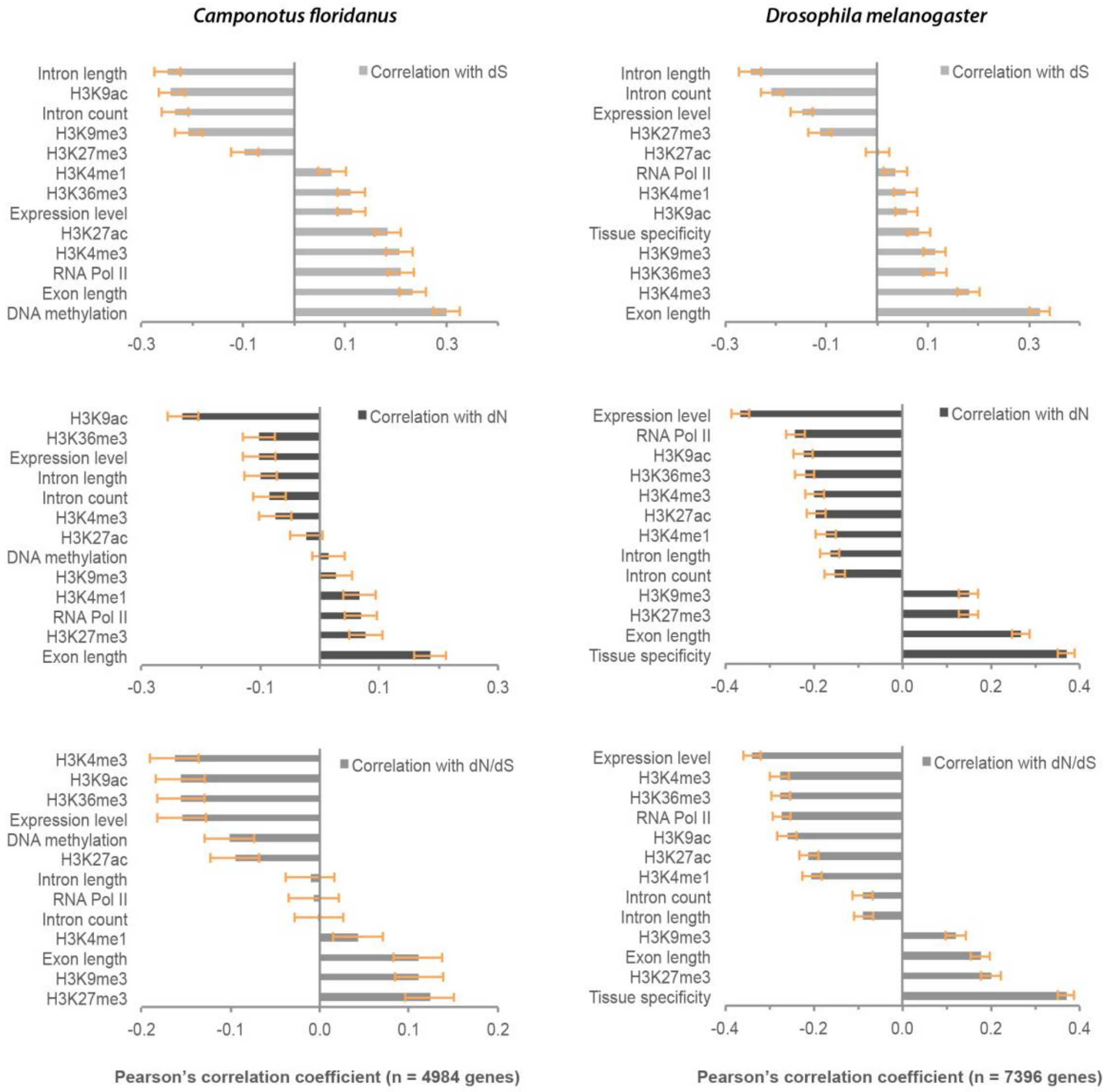


Figure D.1. Relationship between sequence substitution rate and gene characteristics according to Pearson's pairwise correlations in the ant *C. floridanus* and the fly *D. melanogaster*. Correlation coefficients are plotted with 95% confidence intervals.

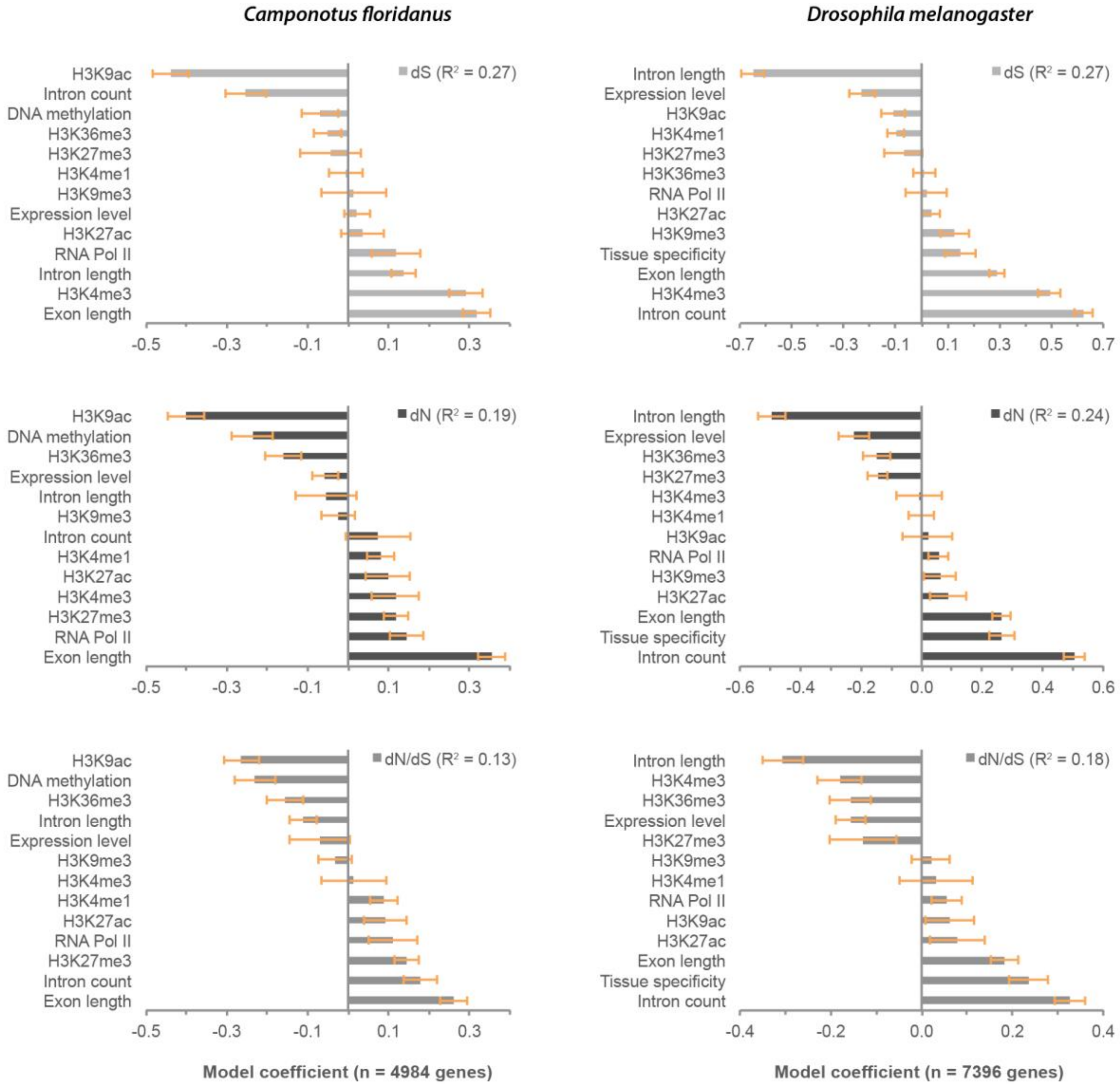


Figure D.2. Relationship between sequence substitution rate and gene characteristics according to multiple linear regression models in the ant *C. floridanus* and the fly *D. melanogaster*. Model coefficients are plotted with 95% confidence intervals.

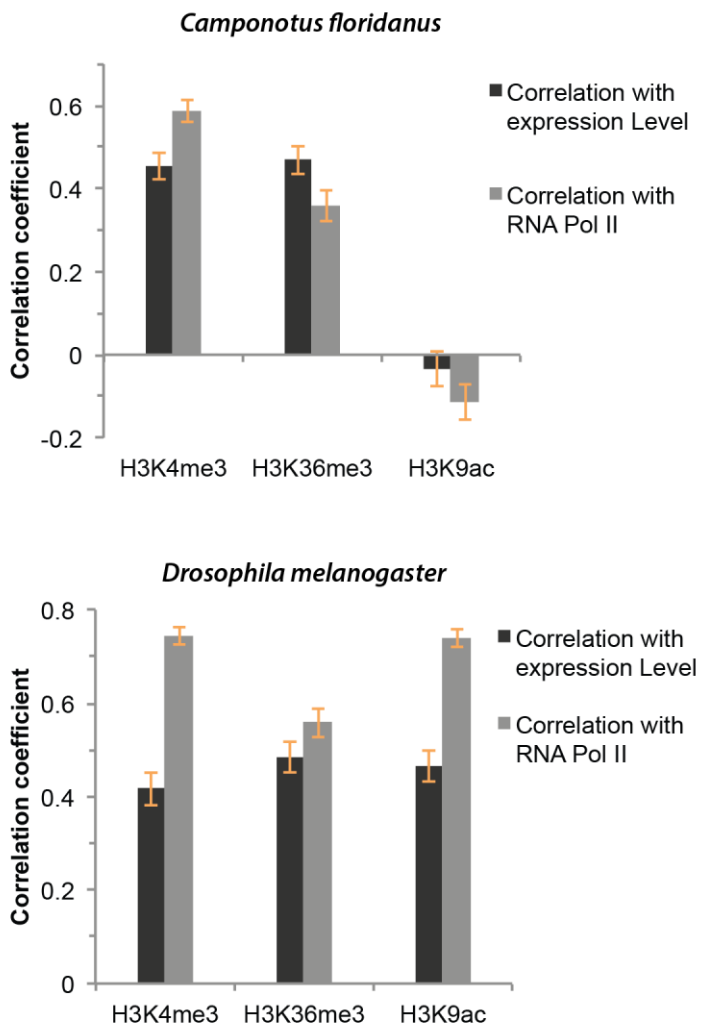


Figure D.3. Correlations between transcriptional activity and the histone modifications H3K4me3, H3K36me3, and H3K9ac. Pearson's correlations with 95% confidence intervals are shown for data from *C. floridanus* and *D. melanogaster* (n = 2102 common ortholog groups).

REFERENCES

Adams, K. L. and J. F. Wendel. 2005. Novel patterns of gene expression in polyploid plants. Trends Genet. 21: 539-543.

Aron, S., L. de Menten, et al. 2005. When hymenopteran males reinvented diploidy. Curr. Biol. 15: 824-827.

Badeaux, A. I. and Y. Shi. 2013. Emerging roles for chromatin as a signal integration and storage platform. Nat. Rev. Mol. Cell Biol. 14: 211-224.

Balbin, O. A., R. Malik, et al. 2015. The landscape of antisense gene expression in human cancers. Genome Res. 25: 1068-1079.

Bannister, A. J. and T. Kouzarides. 2011. Regulation of chromatin by histone modifications. Cell Res. 21: 381-395.

Bell, O., V. K. Tiwari, et al. 2011. Determinants and dynamics of genome accessibility. Nat. Rev. Genet. 12: 554-564.

Benjamini, Y. and Y. Hochberg. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. B 57: 289-300.

Berger, S. L., T. Kouzarides, et al. 2009. An operational definition of epigenetics. Genes Dev. 23: 781-783.

Bergman, C. M., J. W. Carlson, et al. 2005. Drosophila DNase I footprint database: a systematic genome annotation of transcription factor binding sites in the fruitfly, Drosophila melanogaster. Bioinformatics 21: 1747-1749.

Bintu, L., T. Ishibashi, et al. Nucleosomal Elements that Control the Topography of the Barrier to Transcription. Cell 151: 738-749.

Bird, A. P. 1980. DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res. 8: 1499-1504.

Bird, A. P. and A. P. Wolffe. 1999. Methylation-Induced Repression— Belts, Braces, and Chromatin. Cell 99: 451-454.

Bolger, A. M., M. Lohse, et al. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114-2120.

Bonasio, R., Q. Li, et al. 2012. Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and *Harpegnathos saltator*. Curr. Biol. 22: 1755-1764.

Bonasio, R., S. Tu, et al. 2010. Molecular Signals of Epigenetic States. Science 330: 612-616.

Bonasio, R., G. Zhang, et al. 2010. Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. Science 329: 1068-1071.

Brookes, E. and A. Pombo. 2009. Modifications of RNA polymerase II are pivotal in regulating gene expression states. EMBO Rep. 10: 1213-1219.

Burdge, G. C. and K. A. Lillycrop. 2010. Nutrition, epigenetics, and developmental plasticity: implications for understanding human disease. Annu. Rev. Nutr. 30: 315-339.

Carmel, L. and E. V. Koonin. 2009. A universal nonmonotonic relationship between gene compactness and expression levels in multicellular eukaryotes. Genome Biol. Evol. 1: 382-390.

Carmel, L., I. B. Rogozin, et al. 2007. Evolutionarily conserved genes preferentially accumulate introns. Genome Res. 17: 1045-1050.

Cedar, H. and Y. Bergman. 2009. Linking DNA methylation and histone modification: patterns and paradigms. Nat. Rev. Genet. 10: 295-304.

Celniker, S. E., L. A. L. Dillon, et al. 2009. Unlocking the secrets of the genome. Nature 459: 927-930.

Cernilogar, F. M., M. C. Onorati, et al. 2011. Chromatin-associated RNA interference components contribute to transcriptional regulation in Drosophila. Nature 480: 391-395.

Cheng, J., R. Blum, et al. 2014. A role for H3K4 monomethylation in gene repression and partitioning of chromatin readers. Mol. Cell 53: 979-992.

Cheong, J., Y. Yamada, et al. 2006. Diverse DNA methylation statuses at alternative promoters of human genes in various tissues. DNA Res 13: 155-167.

Chodavarapu, R. K., S. Feng, et al. 2010. Relationship between nucleosome positioning and DNA methylation. Nature 466: 388-392.

Chuang, T.-J. and T.-W. Chiang. 2014. Impacts of pretranscriptional DNA methylation, transcriptional transcription factor, and posttranscriptional microRNA regulations on protein evolutionary rate. Genome Biol. Evol. 6: 1530-1541.

Clark, A. G., M. B. Eisen, et al. 2007. Evolution of genes and genomes on the Drosophila phylogeny. Nature 450: 203-218.

Coleman-Derr, D. and D. Zilberman. 2012. Deposition of Histone Variant H2A.Z within Gene Bodies Regulates Responsive Genes. PLoS Genet 8: e1002988.

Conesa, A., S. Gotz, et al. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21: 3674-3676.

Conrad, T. and A. Akhtar. 2012. Dosage compensation in Drosophila melanogaster: epigenetic fine-tuning of chromosome-wide transcription. Nat. Rev. Genet. 13: 123-134.

Deaton, A. M. and A. Bird. 2011. CpG islands and the regulation of transcription. Genes Dev. 25: 1010-1022.

Drewell, R. A., E. C. Bush, et al. 2014. The dynamic DNA methylation cycle from egg to sperm in the honey bee Apis mellifera. Development 141: 2702-2711.

Drummond, D. A., A. Raval, et al. 2006. A single determinant dominates the rate of yeast protein evolution. Mol. Biol. Evol. 23: 327-337.

Drummond, D. A. and C. O. Wilke. 2008. Mistranslation-Induced Protein Misfolding as a Dominant Constraint on Coding-Sequence Evolution. Cell 134: 341-352.

Duret, L. and D. Mouchiroud. 2000. Determinants of Substitution Rates in Mammalian Genes: Expression Pattern Affects Selection Intensity but Not Mutation Rate. Mol. Biol. Evol. 17: 68-070.

Ecker, J. R. and R. W. Davis. 1986. Inhibition of gene expression in plant cells by expression of antisense RNA. Proc Natl Acad Sci USA 83: 5372-5376.

Edgar, B. A. and T. L. Orr-Weaver. 2001. Endoreplication cell cycles: more for less. Cell 105: 297-306.

Eggleton, P. 2011. An Introduction to termites: biology, taxonomy and functional morphology. Biology of Termites: A Modern Synthesis. D. E. Bignell, Y. Roisin and N. Lo. London, Springer**:** 1-27.

Eisenberg, E. and E. Y. Levanon. 2003. Human housekeeping genes are compact. Trends Genet. 19: 362-365.

Elango, N., S. H. Kim, et al. 2008. Mutations of different molecular origins exhibit contrasting patterns of regional substitution rate variation. PLoS Comput. Biol. 4: e1000015.

Engström, P. G., S. J. Ho Sui, et al. 2007. Genomic regulatory blocks underlie extensive microsynteny conservation in insects. Genome Res. 17: 1898-1908.

Evans, J., D. C. A. Shearman, et al. 2004. Molecular basis of sex determination in haplodiploids. Trends in ecology and evolution 19: 1-3.

Evans, J. D., D. C. A. Shearman, et al. 2004. Molecular basis of sex determination in haplodiploids. Trends Ecol. Evol. 19: 1-3.

Evans, J. D. and D. E. Wheeler. 2001. Gene expression and the evolution of insect polyphenisms. BioEssays 23: 62-68.

Feng, S. H., S. J. Cokus, et al. 2010. Conservation and divergence of methylation patterning in plants and animals. Proc. Natl. Acad. Sci. U.S.A. 107: 8689-8694.

Filion, G. J., J. G. van Bemmel, et al. 2010. Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. Cell 143: 212-224.

Flores, K., F. Wolschin, et al. 2012. Genome-wide association between DNA methylation and alternative splicing in an invertebrate. BMC Genomics 13: 480.

Foret, S., R. Kucharski, et al. 2012. DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. Proc. Natl. Acad. Sci. U.S.A. 109: 4968-4973.

Foret, S., R. Kucharski, et al. 2012. DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. Proc Natl Acad Sci USA 109: 4968-4973.

Foret, S., R. Kucharski, et al. 2009. Epigenetic regulation of the honey bee transcriptome: unravelling the nature of methylated genes. BMC Genomics 10: 472.

Fraga, M. F., E. Ballestar, et al. 2005. Epigenetic differences arise during the lifetime of monozygotic twins. Proc. Natl. Acad. Sci. U.S.A. 102: 10604-10609.

Frith, M. C., Y. Fu, et al. 2004. Detection of functional DNA motifs via statistical over-representation. Nucleic Acids Res. 32: 1372-1381.

Galitski, T., A. J. Saldanha, et al. 1999. Ploidy Regulation of Gene Expression. Science 285: 251-254.

Gaunt, M. W. and M. A. Miles. 2002. An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. Mol. Biol. Evol. 19: 748-761.

Gelbart, M. E. and M. I. Kuroda. 2009. Drosophila dosage compensation: a complex voyage to the X chromosome. Development 136: 1399-1410.

Glastad, K. M., M. A. D. Goodisman, et al. 2015. Effects of DNA methylation and chromatin state on rates of molecular evolution in insects. G3.

Glastad, K. M., B. G. Hunt, et al. 2013. Evidence of a conserved functional role for DNA methylation in termites. Insect Mol. Biol. 22: 143-154.

Glastad, K. M., B. G. Hunt, et al. 2015. DNA Methylation and Chromatin Organization in Insects: Insights from the Ant Camponotus floridanus. Genome Biol. Evol. 7: 931-942.

Glastad, K. M., B. G. Hunt, et al. 2011. DNA methylation in insects: on the brink of the epigenomic era. Insect Mol. Biol. 20: 553-565.

Glastad, K. M., B. G. Hunt, et al. 2014. Epigenetic inheritance and genome regulation: is DNA methylation linked to ploidy in haplodiploid insects? Proceedings of the Royal Society B: Biological Sciences 281.

Glastad, K. M., J. Liebig, et al. in preparation. The caste- and sex-specific DNA methylome of the termite *Zootermopsis nevadensis*.

Grant, C. E., T. L. Bailey, et al. 2011. FIMO: scanning for occurrences of a given motif. Bioinformatics 27: 1017-1018.

Guindon, S., J.-F. Dufayard, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59: 307-321.

He, X.-J., T. Chen, et al. 2011. Regulation and function of DNA methylation in plants and animals. Cell Res. 21: 442-465.

Heimpel, G. E. and J. G. d. Boer. 2008. Sex Determination in the Hymenoptera. Annu. Rev. Entomol. 53: 209-230.

Heimpel, G. E. and J. G. de Boer. 2008. Sex determination in the Hymenoptera. Annu. Rev. Entomol. 53: 209-230.

Henikoff, S. 2008. Nucleosome destabilization in the epigenetic regulation of gene expression. Nat. Rev. Genet. 9: 15-26.

Herb, B. R., F. Wolschin, et al. 2012. Reversible switching between epigenetic states in honeybee behavioral subcastes. Nat. Neurosci. 15: 1371-1373.

Hodges, E., A. Smith, et al. 2009. High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing. Genome Res. 19: 1593-1605.

Huh, I., J. Zeng, et al. 2013. DNA methylation and transcriptional noise. Epigenetics & Chromatin 6: 9.

Hunt, B. G., J. A. Brisson, et al. 2010. Functional conservation of DNA methylation in the pea aphid and the honeybee. Genome Biol. Evol. 2: 719-728.

Hunt, B. G., K. M. Glastad, et al. 2013. The function of intragenic DNA methylation: insights from insect epigenomes. Integrative and Comparative Biology 53: 319-328.

Hunt, B. G., K. M. Glastad, et al. 2013. Patterning and regulatory associations of DNA methylation are mirrored by histone modifications in insects. Genome Biol. Evol. 5: 591-598.

Hunt, B. G., L. Ometto, et al. 2013. Evolution at Two Levels in Fire Ants: The Relationship between Patterns of Gene Expression and Protein Sequence Evolution. Mol. Biol. Evol. 30: 263-271.

Hunt, B. G., L. Ometto, et al. 2011. Relaxed selection is a precursor to the evolution of phenotypic plasticity. Proc. Natl. Acad. Sci. U.S.A. 108: 15936-15941.

Jeong, S., G. Liang, et al. 2009. Selective anchoring of DNA methyltransferases 3A and 3B to nucleosomes containing methylated DNA. Mol. Cell. Biol. 29: 5366-5376.

Jones, P. A. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. 13: 484-492.

Kharchenko, P. V., A. A. Alekseyenko, et al. 2011. Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. Nature 471: 480-485.

Klose, R. J. and A. P. Bird. 2006. Genomic DNA methylation: the mark and its mediators. Trends Biochem. Sci. 31: 89-97.

Kolasinska-Zwierz, P., T. Down, et al. 2009. Differential chromatin marking of introns and expressed exons by H3K36me3. Nat. Genet. 41: 376-381.

Koonin, E. V. and Y. I. Wolf. 2010. Constraints and plasticity in genome and molecular-phenome evolution. Nat. Rev. Genet. 11: 487-498.

Kota, S. K. and R. Feil. 2010. Epigenetic transitions in germ cell development and meiosis. Dev. Cell 19: 675-686.

Kozomara, A. and S. Griffiths-Jones. 2014. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 42: D68-D73.

Krieger, M. J. B., K. G. Ross, et al. 1999. Frequency and origin of triploidy in the fire ant Solenopsis invicta. Heredity 82: 142-150.

Krueger, F. and S. R. Andrews. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 27: 1571-1572.

Kucharski, R., J. Maleszka, et al. 2008. Nutritional control of reproductive status in honeybees via DNA methylation. Science 319: 1827-1830.

Kulakovskiy, I. and V. Makeev. 2009. Discovery of DNA motifs recognized by transcription factors through integration of different experimental sources. Biophysics 54: 667-674.

Langley, S. A., G. H. Karpen, et al. 2014. Nucleosomes Shape DNA Polymorphism and Divergence. PLoS Genet 10: e1004457.

Langmead, B., C. Trapnell, et al. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10: R25.

Lavine, L., H. Gotoh, et al. 2015. Exaggerated Trait Growth in Insects. Annu. Rev. Entomol. 60: 453-472.

Lawrie, D. S., P. W. Messer, et al. 2013. Strong purifying selection at synonymous sites in D. melanogaster.

Li-Byarlay, H., Y. Li, et al. 2013. RNA interference knockdown of DNA methyl-transferase 3 affects gene alternative splicing in the honey bee. Proc Natl Acad Sci USA 110: 12750-12755.

Li, H., B. Handsaker, et al. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-2079.

Li, L.-C., S. T. Okino, et al. 2006. Small dsRNAs induce transcriptional activation in human cells. Proc Natl Acad Sci USA 103: 17337-17342.

Li, X. Y., X. F. Wang, et al. 2008. High-resolution mapping of epigenetic modifications of the rice genome uncovers interplay between DNA methylation, histone methylation, and gene expression. Plant Cell 20: 259-276.

Liao, B.-Y., N. M. Scott, et al. 2006. Impacts of Gene Essentiality, Expression Pattern, and Gene Compactness on the Evolutionary Rate of Mammalian Proteins. Mol. Biol. Evol. 23: 2072-2080.

Lo, N., B. Li, et al. 2012. DNA methylation in the termite *Coptotermes lacteus*. Insectes Soc. 59: 257-261.

Lorincz, M. C., D. R. Dickerson, et al. 2004. Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. Nat. Struct. Mol. Biol. 11: 1068-1075.

Love, M., W. Huber, et al. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15: 550.

Löytynoja, A. and N. Goldman. 2005. An algorithm for progressive multiple alignment of sequences with insertions. Proc. Natl. Acad. Sci. U.S.A. 102: 10557-10562.

Luco, R. F., M. Allo, et al. 2011. Epigenetics in Alternative Pre-mRNA Splicing. Cell 144: 16-26.

Luco, R. F., Q. Pan, et al. 2010. Regulation of Alternative Splicing by Histone Modifications. Science 327: 996-1000.

Lyko, F., S. Foret, et al. 2010. The honey bee epigenomes: differential methylation of brain DNA in queens and workers. PLoS Biol. 8: e1000506.

Makova, K. D. and R. C. Hardison. 2015. The effects of chromatin organization on variation in mutation rates in the genome. Nat. Rev. Genet. 16: 213-223.

Mao, L., G. Henderson, et al. 2005. Formosan subterranean termite (Isoptera: *Rhinotermitidae*) soldiers regulate juvenile hormone levels and caste differentiation in workers. Ann. Entomol. Soc. Am. 98: 340-345.

Margueron, R. and D. Reinberg. 2010. Chromatin structure and the inheritance of epigenetic information. Nat. Rev. Genet. 11: 285-296.

Maunakea, A. K., I. Chepelev, et al. 2013. Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. Cell Res. 23: 1256-1269.

Maunakea, A. K., R. P. Nagarajan, et al. 2010. Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature 466: 253-257.

McLeay, R. and T. Bailey. 2010. Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. BMC Bioinformatics 11: 165.

Miura, T. and M. E. Scharf. 2011. Biology of Termites: A Modern Synthesis. London, Springer.

Mugal, C. F. and H. Ellegren. 2011. Substitution rate variation at human CpG sites correlates with non-CpG divergence, methylation level and GC content. Genome Biol 12: R58.

Nanty, L., G. Carbajosa, et al. 2011. Comparative methylomics reveals gene-body H3K36me3 in *Drosophila* predicts DNA methylation and CpG landscapes in other invertebrates. Genome Res. 21: 1841-1850.

Negre, N., C. D. Brown, et al. 2011. A cis-regulatory map of the Drosophila genome. Nature 471: 527-531.

Nipitwattanaphon, M., J. Wang, et al. Forthcoming. Effects of ploidy and sex-locus genotype on gene expression patterns in the fire ant *Solenopsis invicta*.

Okitsu, C. Y. and C. L. Hsieh. 2007. DNA methylation dictates histone H3K4 methylation. Mol. Cell. Biol. 27: 2746-2757.

Ometto, L., D. Shoemaker, et al. 2011. Evolution of gene expression in fire ants: the effects of developmental stage, caste, and species. Mol. Biol. Evol. 28: 1381-1392.

Ooi, S. K. T., C. Qiu, et al. 2007. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 448: 714-717.

Otto, S. P. 2007. The Evolutionary Consequences of Polyploidy. Cell 131: 452-462.

Otto, S. P. and P. Jarne. 2001. Haploids--Hapless or Happening? Science 292: 2441-2443.

Pal, C., B. Papp, et al. 2006. An integrated view of protein evolution. Nat Rev Genet 7: 337-348.

Park, C., W. Qian, et al. 2012. Genomic evidence for elevated mutation rates in highly expressed genes. EMBO Rep. 13: 1123-1129.

Park, P. J. 2009. ChIP-seq: advantages and challenges of a maturing technology. Nat. Rev. Genet. 10: 669-680.

Park, Y., M. E. Figueroa, et al. 2014. MethylSig: a whole genome DNA methylation analysis pipeline. Bioinformatics 30: 2414-2422.

Payer, B. and J. T. Lee. 2008. X Chromosome Dosage Compensation: How Mammals Keep the Balance. Annu. Rev. Genet. 42: 733-772.

Penn, O., E. Privman, et al. 2010. GUIDANCE: a web server for assessing alignment confidence scores. Nucleic Acids Res. 38: W23-W28.

Pfennig, D. W., M. A. Wund, et al. 2010. Phenotypic plasticity's impacts on diversification and speciation. Trends Ecol. Evol. 25: 459-467.

Prendergast, J., H. Campbell, et al. 2007. Chromatin structure and evolution in the human genome. BMC Evol. Biol. 7: 72.

Prendergast, J. G. D. and C. A. M. Semple. 2011. Widespread signatures of recent selection linked to nucleosome positioning in the human lineage. Genome Res. 21: 1777-1787.

Qu, W., S.-i. Hashimoto, et al. 2012. Genome-wide genetic variations are highly correlated with proximal DNA methylation patterns. Genome Res. 22: 1419-1425.

R Development Core Team. 2011. R: a language and environment for statistical computing. Vienna, Austria, R Foundation for Statistical Computing.

R Development Core Team. 2013. R: a language and environment for statistical computing. Vienna, Austria, R Foundation for Statistical Computing.

Rasch, E., J. Cassidy, et al. 1977. Evidence for dosage compensation in parthenogenetic Hymenoptera. Chromosoma 59: 323-340.

Reik, W. 2007. Stability and flexibility of epigenetic gene regulation in mammalian development. Nature 447: 425-432.

Roberts, A., H. Pimentel, et al. 2011. Identification of novel transcripts in annotated genomes using RNA-Seq. Bioinformatics 27: 2325-2329.

Roberts, A., C. Trapnell, et al. 2011. Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biol. 12: R22.

Robinson, G. E., C. M. Grozinger, et al. 2005. Sociogenomics: social life in molecular terms. Nat Rev Genet 6: 257-270.

Ross, K. G. and D. J. C. Fletcher. 1985. Genetic origin of male diploidy in the fire ant, *Solenopsis invicta* (Hymenoptera: Formicidae), and its evolutionary significance. Evolution 39: 888-903.

Ross, K. G., E. L. Vargo, et al. 1993. Effect of a founder event on variation in the genetic sex-determining system of the fire ant *Solenopsis invicta*. Genetics 135: 843-854.

Ross, K. G., E. L. Vargo, et al. 1993. Effect of a founder event on variation in the genetic sex-determining system of the fire ant Solenopsis invicta. Genetics 135: 843-854.

Sarda, S., J. Zeng, et al. 2012. The evolution of invertebrate gene body methylation. Mol. Biol. Evol. 29: 1907-1916.

Sassone-Corsi, P. 2002. Unique Chromatin Remodeling and Transcriptional Regulation in Spermatogenesis. Science 296: 2176-2178.

Scharf, M. E., C. E. Buckspan, et al. 2007. Regulation of polyphenic caste differentiation in the termite Reticulitermes flavipes by interaction of intrinsic and extrinsic factors. J. Exp. Biol. 210: 4390-4398.

Schmitz, R. J. and J. R. Ecker. 2012. Epigenetic and epigenomic variation in *Arabidopsis thaliana*. Trends Plant Sci. 17: 149-154.

Scholes, D. R., A. V. Suarez, et al. 2013. Can endopolyploidy explain body size variation within and between castes in ants? Ecology and Evolution 3: 2128-2137.

Schuster-Bockler, B. and B. Lehner. 2012. Chromatin organization is a major influence on regional mutation rates in human cancer cells. Nature 488: 504-507.

Schwartz, S., E. Meshorer, et al. 2009. Chromatin organization marks exon-intron structure. Nat. Struct. Mol. Biol. 16: 990-U117.

Shao, Z., Y. Zhang, et al. 2012. MAnorm: a robust model for quantitative comparison of ChIP-Seq data sets. Genome Biol. 13: 1-17.

Shenker, N. and J. M. Flanagan. 2012. Intragenic DNA methylation: implications of this epigenetic mechanism for cancer research. Br. J. Cancer 106: 248-253.

Shukla, S., E. Kavak, et al. 2011. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature 479: 74-79.

Simola, D. F., L. Wissler, et al. 2013. Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. Genome Res.

Simola, D. F., C. Ye, et al. 2013. A chromatin link to caste identity in the carpenter ant Camponotus floridanus. Genome Res. 23: 486-496.

Smith, C. R., A. L. Toth, et al. 2008. Genetic and genomic analyses of the division of labour in insect societies. Nat. Rev. Genet. 9: 735-748.

Smyth, G. K. 2005. limma: Linear Models for Microarray Data. Bioinformatics and Computational Biology Solutions Using R and Bioconductor. R. Gentleman, V. Carey, W. Huber, R. Irizarry and S. Dudoit, Springer New York**:** 397-420.

Stasevich, T. J., Y. Hayashi-Takanaka, et al. 2014. Regulation of RNA polymerase II activation by histone acetylation in single living cells. Nature 516: 272-275.

Suzuki, M. M. and A. Bird. 2008. DNA methylation landscapes: provocative insights from epigenomics. Nat. Rev. Genet. 9: 465-476.

Takayama, S., J. Dhahbi, et al. 2014. Genome methylation in D. melanogaster is found at specific short motifs and is independent of DNMT2 activity. Genome Res. 24: 821-830.

Takuno, S. and B. S. Gaut. 2012. Body-methylated genes in Arabidopsis thaliana are functionally important and evolve slowly. Mol. Biol. Evol. 29: 219-227.

Talavera, G. and J. Castresana. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56: 564-577.

Talia, S. D., J. M. Skotheim, et al. 2007. The effects of molecular noise and size control on variability in the budding yeast cell cycle. Nature 448: 947-951.

Tan, Y., B. Zhang, et al. 2009. Transcriptional inhibiton of Hoxd4 expression by miRNA-10a in human breast cancer cells. BMC Mol. Biol. 10: 12.

Terrapon, N., C. Li, et al. 2014. Molecular traces of alternative social organization in a termite genome. Nat Commun 5.

Tolstorukov, M. Y., N. Volfovsky, et al. 2011. Impact of chromatin structure on sequence variability in the human genome. Nat. Struct. Mol. Biol. 18: 510-515.

Toru, M. and M. E. Scharf. 2011. Molecular basis underlying caste differentiation in termites. Biology of Termites: A Modern Synthesis. D. E. Bignell, Y. Roisin and N. Lo. London, Springer**:** 211-253.

Townsend, J. and D. Hartl. 2002. Bayesian analysis of gene expression levels: statistical quantification of relative mRNA level across multiple strains or treatments. Genome Biol. 3: research0071.

Townsend, J. P. and D. L. Hartl. 2002. Bayesian analysis of gene expression levels: statistical quantification of relative mRNA level across multiple strains or treatments. Genome Biol. 3.

Trapnell, C., L. Pachter, et al. 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25: 1105-1111.

Tufarelli, C., J. A. S. Stanley, et al. 2003. Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. Nat. Genet. 34: 157-165.

Urieli-Shoval, S., Y. Gruenbaum, et al. 1982. The absence of detectable methylated bases in *Drosophila melanogaster* DNA. FEBS Lett. 146: 148-152.

Venkatesh, S., M. Smolle, et al. 2012. Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. Nature 489: 452-455.

Wall, D. P., A. E. Hirsh, et al. 2005. Functional genomic analysis of the rates of protein evolution. Proc. Natl. Acad. Sci. U.S.A. 102: 5483-5488.

Wang, H., M. T. Maurano, et al. 2012. Widespread plasticity in CTCF occupancy linked to DNA methylation. Genome Res. 22: 1680-1688.

Wang, J., S. Jemielity, et al. 2007. An annotated cDNA library and microarray for large-scale gene-expression studies in the ant Solenopsis invicta. Genome Biol. 8.

Wang, J., S. Jemielity, et al. 2007. An annotated cDNA library and microarray for large-scale gene-expression studies in the ant *Solenopsis invicta*. Genome Biol. 8: R9.

Wang, J., Y. Wurm, et al. 2013. A Y-like social chromosome causes alternative colony organization in fire ants. Nature 493: 664-668.

Wang, X., X. Fang, et al. 2014. The locust genome provides insight into swarm formation and long-distance flight. Nature communications 5.

Wang, X., D. Wheeler, et al. 2013. Function and Evolution of DNA Methylation in *Nasonia vitripennis*. PLoS Genet 9: e1003872.

Wang, Z. and J. Zhang. 2011. Impact of gene expression noise on organismal fitness and the efficacy of natural selection. Proc Natl Acad Sci USA 108: E67–E76.

Wang, Z. and J. Zhang. 2011. Impact of gene expression noise on organismal fitness and the efficacy of natural selection. Proc. Natl. Acad. Sci. U.S.A. 108: E67–E76.

Waterhouse, R. M., E. M. Zdobnov, et al. 2011. OrthoDB: the hierarchical catalog of eukaryotic orthologs in 2011. Nucleic Acids Res. 39: D283-D288.

Weber, M., I. Hellmann, et al. 2007. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat. Genet. 39: 457-466.

Wedeles, Christopher J., Monica Z. Wu, et al. 2013. Protection of Germline Gene Expression by the C. elegans Argonaute CSR-1. Dev. Cell 27: 664-671.

WEINBERG, M. S., L. M. VILLENEUVE, et al. 2006. The antisense strand of small interfering RNAs directs histone methylation and transcriptional gene silencing in human cells. Rna 12: 256-262.

West-Eberhard, M. J. 2003. Developmental plasticity and evolution. New York, NY, Oxford University Press.

Wheeler, D. E. 1986. Developmental and physiological determinants of caste in social Hymenoptera: evolutionary implications. Am. Nat.: 13-34.

Wiegmann, B., M. Trautwein, et al. 2009. Single-copy nuclear genes resolve the phylogeny of the holometabolous insects. BMC Biology 7: 34.

Wilson, E. O. 1990. Success and dominance in ecosystems: the case of the social insects. Success and dominance in ecosystems: the case of the social insects.

Wolf, Y. I., P. S. Novichkov, et al. 2009. The universal distribution of evolutionary rates of genes and distinct characteristics of eukaryotic genes of different apparent ages. Proc Natl Acad Sci USA 106: 7273-7280.

Wurm, Y., J. Wang, et al. 2011. The genome of the fire ant *Solenopsis invicta*. Proc. Natl. Acad. Sci. U.S.A. 108: 5679-5684.

Yamanaka, N., K. F. Rewitz, et al. 2013. Ecdysone control of developmental transitions: lessons from Drosophila research. Annu. Rev. Entomol. 58: 497.

Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24: 1586-1591.

Yearim, A., S. Gelfman, et al. 2015. HP1 is involved in regulating the global impact of DNA methylation on alternative splicing. Cell reports 10: 1122-1134.

Yi, S. V. and M. A. D. Goodisman. 2009. Computational approaches for understanding the evolution of DNA methylation in animals. Epigenetics 4: 551-556.

Zemach, A., I. E. McDaniel, et al. 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science 328: 916-919.

Zentner, G. E. and S. Henikoff. 2013. Regulation of nucleosome dynamics by histone modifications. Nat. Struct. Mol. Biol. 20: 259-266.

Zhang, X. Y., Y. V. Bernatavichute, et al. 2009. Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*. Genome Biol. 10.

Zhang, Y., T. Liu, et al. 2008. Model-based Analysis of ChIP-Seq (MACS). Genome Biol. 9: R137.

Zhou, V. W., A. Goren, et al. 2011. Charting histone modifications and the functional organization of mammalian genomes. Nat. Rev. Genet. 12: 7-18.

Zilberman, D., D. Coleman-Derr, et al. 2008. Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. Nature 456: 125-129.

Zilberman, D., M. Gehring, et al. 2007. Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. Nat. Genet. 39: 61-69.