


Lineage and Parent-of-Origin Effects in DNA Methylation of Honey Bees (*Apis mellifera*) Revealed by Reciprocal Crosses and Whole-Genome Bisulfite Sequencing

Xin Wu¹, David A. Galbraith^{2,†}, Paramita Chatterjee¹, Hyeonsoo Jeong¹, Christina M. Grozinger², and Soojin V. Yi ^{1,*}

¹School of Biological Sciences, Institute for Bioengineering and Bioscience, Georgia Institute of Technology, Atlanta, Georgia

²Department of Entomology, Center for Pollinator Research, Huck Institutes of the Life Sciences, Pennsylvania State University

*Corresponding author: E-mail: soojinyi@gatech.edu.

Accepted: 22 June 2020

[†]Present address: Bristol-Myers Squibb, Lawrence, NJ

Abstract

Parent-of-origin methylation arises when the methylation patterns of a particular allele are dependent on the parent it was inherited from. Previous work in honey bees has shown evidence of parent-of-origin-specific expression, yet the mechanisms regulating such pattern remain unknown in honey bees. In mammals and plants, DNA methylation is known to regulate parent-of-origin effects such as genomic imprinting. Here, we utilize genotyping of reciprocal European and Africanized honey bee crosses to study genome-wide allele-specific methylation patterns in sterile and reproductive individuals. Our data confirm the presence of allele-specific methylation in honey bees in lineage-specific contexts but also importantly, though to a lesser degree, parent-of-origin contexts. We show that the majority of allele-specific methylation occurs due to lineage rather than parent-of-origin factors, regardless of the reproductive state. Interestingly, genes affected by allele-specific DNA methylation often exhibit both lineage and parent-of-origin effects, indicating that they are particularly labile in terms of DNA methylation patterns. Additionally, we re-analyzed our previous study on parent-of-origin-specific expression in honey bees and found little association with parent-of-origin-specific methylation. These results indicate strong genetic background effects on allelic DNA methylation and suggest that although parent-of-origin effects are manifested in both DNA methylation and gene expression, they are not directly associated with each other.

Key words: epigenetics, intragenomic conflict, honey bees, social insects, DNA methylation.

Significance

Preferential expression of one parental allele over the other has been long observed in mammals and plants, and recently, in honey bees. Complexities of honey bee societies may provide a fertile ground for parent-of-origin-specific gene expression to resolve intragenomic conflict. Here, we examined DNA methylation in honey bees as a potential molecular mechanism for driving parent-of-origin expression, using whole-genome bisulfite sequencing of reciprocal crosses between genetically divergent strains. We present data consistent with parent-of-origin-specific DNA methylation; however, we did not detect a direct link between parent-of-origin-specific expression and methylation. Moreover, the strongest factor of allele-specific methylation was the effect of genetic background. Our findings begin to resolve the complex relationship between allele-specific methylation and expression in insect genomes and the factors driving these phenomena.

© The Author(s) 2020. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Introduction

Parent-specific expression occurs when the expression of an allele varies depending on whether it was inherited from the mother (matrigene) or father (patrigene) and can even be associated with complete silencing of expression of one of the parental alleles (Haig 2000; Wilkins and Haig 2003; Wolf and Hager 2006; Ferguson-Smith 2011). Several theories have been proposed to explain the causes and evolutionary origins of parent-specific expression (e.g., Patten et al. 2014). Among them, Haig's kinship theory of intragenomic conflict hypothesizes that parent-specific expression is due to maternally and paternally inherited genes being under different selection pressures (Haig 2000; Pegoraro et al. 2017). In cases where a female produces offspring with multiple males, selection may favor matrigene expression profiles that promote equal distribution of resources among siblings, whereas patrigene expression profiles would support processes that prioritize individual fitness (Haig 2000; Pegoraro et al. 2017). Support for this theory has been provided in mammalian and plant species, where matrigene expression tends to be associated with impeded growth, whereas patrigene expression stimulates it (Haig 2000; Wilkins and Haig 2003). The kinship theory of intragenomic conflict is applicable to social insects because the myriad of differences in matrigene and patrigene relatedness found in individuals in social insect colonies (Queller 2003). Honey bees are a particularly dramatic example in which intragenomic conflict is thought to play an important role in shaping social behaviors, because honey bees are haplodiploid, and the single reproductive female queen mates with ten or more males, giving rise to offspring with significant variation in their matrigene–patrigene relatedness (Queller 2003; Kocher et al. 2015; Galbraith, Kocher, et al. 2016; Pegoraro et al. 2017).

Transcriptomic studies in social insects have provided support for the kinship theory of intragenomic conflict (Bonasio et al. 2012; Lonsdale et al. 2017). Kocher et al. (2015) showed largely consistent patterns of parent-specific expression across developmental stages, behavioral states, and tissues, though these findings may have been confounded by lineage effects (Gibson et al. 2015). The kinship theory of intragenomic conflict predicts that patrigenes should stimulate worker reproduction during queenless conditions (Haig 1992; Queller 2003). Oldroyd et al. (2014) and Galbraith, Kocher, et al. (2016) demonstrated that worker ovary size (a trait that is positively correlated with worker ovary activation rates under queenless conditions) and worker ovary activation are correlated with paternal genotype in two different studies (Oldroyd et al. 2014; Galbraith, Kocher, et al. 2016). Galbraith, Kocher, et al. (2016) further demonstrated that patrigenes were preferentially upregulated in reproductive tissues of reproductive workers versus sterile workers in reciprocal crosses between Africanized and European honey bee stocks.

Although these observations support intragenomic conflict and parent-of-origin effects, the identities of molecular mechanisms behind these phenomena in social insects remain unclear (Kocher et al. 2015; Galbraith, Kocher, et al. 2016; Galbraith, Yi, et al. 2016). In mammalian and plant species, epigenetic regulation primarily in the form of DNA methylation is thought to mediate parent-specific expression and parent-specific silencing of alleles (Reik and Walter 2001; Bird 2002; Queller 2003; Law and Jacobsen 2010). Intriguingly, the honey bee also possesses a fully functional DNA methylation system (Wang et al. 2006; Lyko et al. 2010). However, unlike in mammalian and plant species, DNA methylation in honey bees is found primarily in gene bodies and coding regions, and thus may not have the same function (Elango et al. 2009; Lyko et al. 2010; Zeng and Yi 2010; Galbraith et al. 2015). Gene body methylation has been shown to be positively correlated with gene expression in insects (Lyko et al. 2010; Zeng and Yi 2010; Wang et al. 2013; Galbraith et al. 2015; Wu et al. 2020), which would suggest that allele-specific methylation and expression would also share the same direction of association. A previous study in the hymenopteran wasp *Nasonia vitripennis* utilizing reciprocal crosses found strong lineage-specific effects where the hybrid offspring recapitulated the methylation levels of nearly all of the differentially methylated genes (DMGs) between their parents (Wang et al. 2016). Despite the near complete transmission of lineage-specific methylation patterns to the hybrid offspring and strong correlations between allele-specific methylation and expression, there was no evidence of parent-of-origin effects on DNA methylation (Wang et al. 2016).

Here, we utilized samples collected from a previous study of reciprocal crosses between two divergent strains of honey bees (Galbraith, Kocher, et al. 2016) to survey genome-wide DNA methylation using whole-genome bisulfite sequencing (WGBS). For these studies, sterile and reproductive honey bee workers were generated from reciprocal crosses of European and Africanized honey bee stocks. These samples allow us to evaluate allele-specific DNA methylation patterns across the genome and determine if these are associated with parent, lineage, or reproductive state differences, taking advantage of the lineage-based single-nucleotide polymorphisms (SNPs) present in these two stocks. By uncoupling of parent and lineage effects, and pairing nucleotide-resolution DNA methylation data with gene expression data from the same experiment, we uncover underlying genetic and parent-specific effects on DNA methylation and determine if these are associated with behavioral and physiological variation in worker honey bees.

Materials and Methods

Biological Sample Collection

Rearing and collection of honey bee samples were described in Galbraith, Kocher, et al. (2016). We obtained a total of

eight sterile and eight reproductive workers that were equally sourced from genetic blocks A and B and European_{mother} × Africanized_{father} and Africanized_{mother} × European_{father} crosses (one of each cross per block, and thus two sterile and two reproductive workers from each cross). Our samples were collected from the same colonies as those used for transcriptomic analysis in Galbraith, Kocher, et al. (2016) and were the full-sisters of the workers used in that study (same mother and same father). Genomic DNA was extracted from the reproductive tissues (ovaries and abdominal fat bodies) of each individual, as in Galbraith, Kocher, et al. (2016). The DNA was collected from each individual for bisulfite sequencing library preparation using a Gentra Puregene Tissue kit (Qiagen, Cat. No. 158667).

Whole-Genome Bisulfite Library Preparation and Sequencing

The libraries were made with an Illumina sequencer compatible protocol (Urich et al. 2015). The extracted DNA was fragmented using the Covaris S-series focused ultrasonicator (Covaris Cat. No. S220) using the “500-bp-target peak size” protocol. The fragmented DNA was then size selected (200–600 bp) with the SPRI bead-based size selection protocol (Urich et al. 2015). The DNA end repair step was then performed where DNA fragments containing damaged or incompatible 5'- and/or 3'-protruding ends are converted to 5'-phosphorylated, blunt-ended DNA. Blunt-ended DNA fragments were ligated with 3'-dAMP overhangs and methylated adapters.

We added 10 ng of unmethylated lambda phage DNA (Promega, Cat. No. D1501) to the input DNA of each library as a control for bisulfite conversion rates. Bisulfite treatment of genomic DNA was performed using the MethylCode Bisulfite Conversion Kit (Life Technologies, Cat. No. MECOV-50). With this method, nonmethylated cytosines are converted to uracil and read as thymine (T) when sequenced. Methylated cytosines protected from conversion are still read as cytosine (C). Briefly, purified genomic DNA is treated with CT conversion reagent in a thermocycler for 10 min at 98 °C, followed by 2.5 h at 64 °C. Low-cycle (4–8) polymerase chain reaction amplification was performed with Kapa HiFi Uracil Hotstart polymerase enzyme (Kapa Biosystems, Cat. No. KK2801) which can tolerate uracil residues. Libraries were quantified by using Qubit high sensitivity DNA kit (Life Technologies, Cat. No. Q32851) and Agilent High Sensitivity DNA Bioanalyzer (Agilent Technologies, Cat. No. 5067-4626). The libraries were diluted and loaded onto the Illumina HiSeq X at the Macrogen Laboratory for sequencing using 150-bp paired-end reads.

Generating N-Masked Genomes

A list of SNPs for each parent from each cross was obtained from a previous study (Galbraith, Kocher, et al. 2016). For

each cross, we used stringent criteria to generate a list of informative SNPs. We removed ambiguous SNPs (e.g., heterozygous SNPs in the diploid queen genome and SNPs that would be identical in both drone and queen), SNPs with a Phred score of <30, and C → T and T → C SNPs. We also discarded SNPs with a read depth of <5 in either the European or the Africanized alleles. In addition, only conserved SNPs from both of the crosses in each genetic block were retained, which resulted in 213,056 and 214,504 informative SNPs for genetic blocks A and B, respectively (supplementary table 1, Supplementary Material online). Using the resulting list of informative SNPs, we used a custom Python script (doi:10.5281/zenodo.3873279) to generate one N-masked genome for each genetic block.

WGBS Data Processing

Raw reads were trimmed for low quality and adapters using Trim_galore! (Martin 2011) with default parameters. Trimmed reads of each individual were then aligned to their respective N-masked genomes based on genetic block using Bismark and default parameters (Krueger and Andrews 2011). Additionally, reads were aligned to the unmethylated lambda genome (GenBank accession number: J02459.1) to estimate the bisulfite conversion efficiency based on the deamination rate. Following alignment, each aligned read was split into European or Africanized alleles using SNPSplit (Krueger and Andrews 2016) in paired-end mode based on the previously generated list of informative SNPs for each cross (supplementary table 1, Supplementary Material online). Each allele-specific alignment, which we will refer to as one sample, was then deduplicated using Bismark (Krueger and Andrews 2011).

For each sample, we applied the binomial test at each CpG site using the deamination rate as the probability of success and a false discovery rate (FDR) threshold of <0.05 (Benjamini and Hochberg 1995) to classify each CpG site as either “methylated” or “unmethylated” (Lyko et al. 2010; Wang et al. 2013; Galbraith et al. 2015). For downstream analyses, we only retained CpG sites that were classified as “methylated” in at least one sample (Huh et al. 2019).

WGBS reads are uploaded on SRA and accessible under BioProject PRJNA608132.

Differential Methylation Analysis

We used the DSS package (Park and Wu 2016) to identify CpGs that were differentially methylated (referred to as differentially methylated positions, or DMPs). DSS handles variance across biological replicates as well as models read counts from WGBS experiments, while accounting for effects of biological covariates. Specifically, we considered the parent-of-origin effect (paternal or maternal) as well as the lineage effect (Africanized or European) and corrected for multiple testing by FDR threshold of <0.1. We performed this analysis

separately for each genetic block. In addition, a significant site was required to exhibit at least 60% relative allele-specific methylation bias in both reciprocal crosses (European_{mother} × Africanized_{father} and Africanized_{mother} × European_{father}), similarly to how allele-specific expression bias was previously calculated (Kocher et al. 2015; Galbraith, Kocher, et al. 2016). The relative allele-specific methylation is calculated as the percent of methylation (Schultz et al. 2012; Galbraith et al. 2015; Lindsey et al. 2018) of one allele relative to the total methylation of both alleles. DMGs for each independent variable were then defined as genes containing at least one DMP that all showed the same direction of bias (Galbraith et al. 2015; Kocher et al. 2015). We classified genes using DMPs for each variable independent of DMPs from other variables, meaning that a gene containing DMPs from multiple variables can be a DMG in multiple categories as long as it satisfied the previous requirement each time (fig. 3).

In addition to analyzing each individual CpG, we assessed allele-specific methylation at the gene body level using a generalized linear model of the binomial family similar to previous studies (Lyko et al. 2010; Galbraith et al. 2015). Briefly, the number of methylated and unmethylated reads was used as the response vector and modeled by three categorical variables: lineage (Africanized or European), parent (Paternal or Maternal), and CpG position (Lyko et al. 2010; Galbraith et al. 2015). Similar to our CpG level analysis, we applied this model to samples for each genetic block and reproductive status separately.

RNA-seq Data Processing

RNA-seq data were previously generated (Galbraith, Kocher, et al. 2016). We reanalyzed these data following the same criteria as in the methylation analysis, so that the results can be directly compared. Reads (accession number: GSE76164) were trimmed using the same parameters as the WGBS data and aligned to each sample's respective N-masked genome using HISAT2 with soft clipping disabled (parameter setting: –sp 1000,1000). As with the WGBS data, SNPSplit was used to separate European and Africanized alleles, whereas HTSeq (Anders et al. 2015) with default parameters was used to count allele-separated reads within coding regions of the genome. The DESeq2 package (Love et al. 2014) was used to identify differentially expressed genes (DEGs). Gene expression was quantified based on normalized counts using the “estimateSizeFactors” function from the DESeq2 package. Differential expression was assessed using the same predictors as our differential methylation analysis (parent and lineage) and the normalized gene counts as the response variable. This analysis was also performed separately for each genetic block. Significance level for genes was at an FDR-adjusted (Benjamini and Hochberg 1995) level of 0.1 and required a 60% relative expression bias per allele similar to the methylation analysis.

Gene Ontology

Gene ontology (GO) analysis was performed by uploading BeeBase gene IDs (Elsik et al. 2014) onto the DAVID bioinformatics Resources 6.8 webpage and using the Functional Annotation tool (Huang da et al. 2009). GO term enrichment was assessed at a significance level of $P < 0.05$. The background gene list for enrichment analyses was all protein-coding genes in the genome.

Results

Lineage and Parent-of-Origin Effects on DNA Methylation

To study allele-specific patterns in DNA methylation, we generated a list of informative SNPs that could be used to reliably differentiate our sequenced reads and assign them to the appropriate parental allele. Because the parents of each genetic block were from different colonies, we assessed variation of DNA methylation for each block separately. Due to the large amount of informative SNPs that are unique to each genetic block, analyzing the genetic blocks separately enabled us to increase the scope of our analysis by surveying a larger portion of the genome as well as giving us an indicator of the robustness of our results (supplementary table 1, Supplementary Material online). Within each genetic block, only informative SNPs that were present in both crosses (European_{mother} × Africanized_{father} and Africanized_{mother} × European_{father}) were used for our analyses. In genetic block A, there were 213,056 such SNPs, which allowed us to assign alleles to 48,745 methylated CpGs that were distributed within 5,613 genes. Similarly, there were 214,504 informative SNPs in genetic block B, allowing us to detect 41,764 methylated CpGs within 5,359 genes. In contrast, the number of informative SNPs that were shared between the two blocks was only 27,146, which led to 2,506 methylated CpGs that could be analyzed, yielding little power (supplementary fig. 1, Supplementary Material online). Consequently, we selected to analyze the two blocks separately to improve our ability to detect differential DNA methylation.

We used a linear model framework to identify individual CpGs that exhibited variation of DNA methylation among

Table 1

Summary of Differentially Methylated Positions Based on Reproductive State and Direction of Allele-Specific Bias

	Block A		Block B	
	Sterile	Reproductive	Sterile	Reproductive
Parent-of-origin				
Maternal bias	132	190	208	189
Paternal bias	148	218	216	188
Lineage				
Africanized bias	333	921	696	727
European bias	410	948	829	964

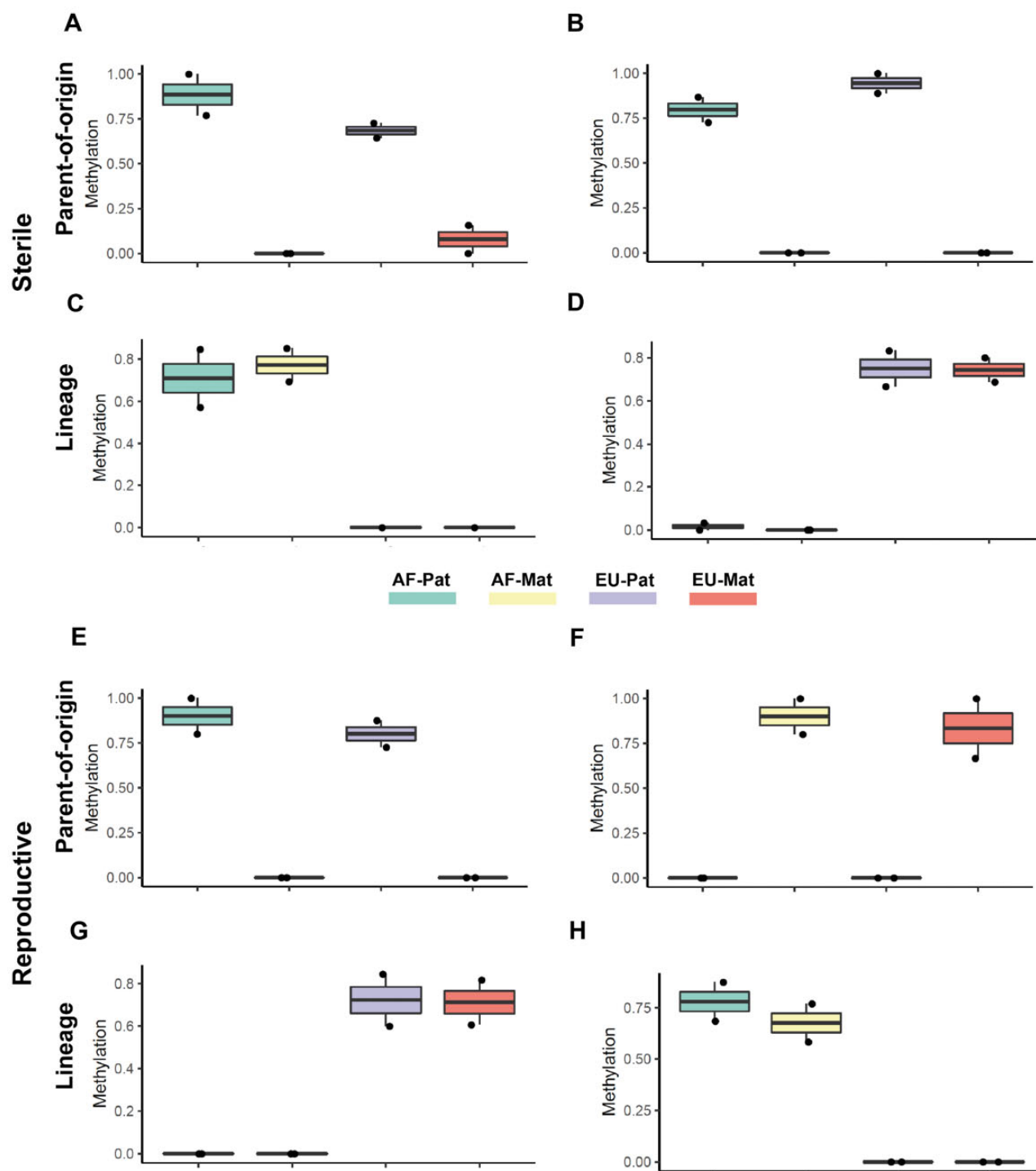


Fig. 1—Examples of allele-specific methylation bias at the CpG level (DMPs). Each data point represents the fractional methylation level of one sample at the position. (A and B) DMPs located on scaffolds Group13.4 and Group3.15 at positions 107256 (gene ID: GB48238) and 474358 (gene ID: GB47021), respectively, exhibiting parent-of-origin bias (paternal) in the sterile workers. (C) An example of an Africanized-biased DMP on scaffold Group8.12 position 54654 (gene ID: GB54805) and (D) a European-biased DMP located on scaffold Group1.7 position 295042 (gene ID: GB53211) in sterile workers. In the reproductive workers, (E) depicts a paternal-biased DMP (scaffold Group9.10 position 1931896; gene ID: GB42829), whereas (F) is a maternal-biased DMP (scaffold Group1.41 position 1013225; gene ID: GB55015). Lineage-biased DMPs in reproductive workers are shown in (G) (scaffold Group 1.14 position 2081; gene ID: GB51836) and (H) (scaffold Group7.12 position 453285; gene ID: GB49222), biased toward Africanized and European workers, respectively.

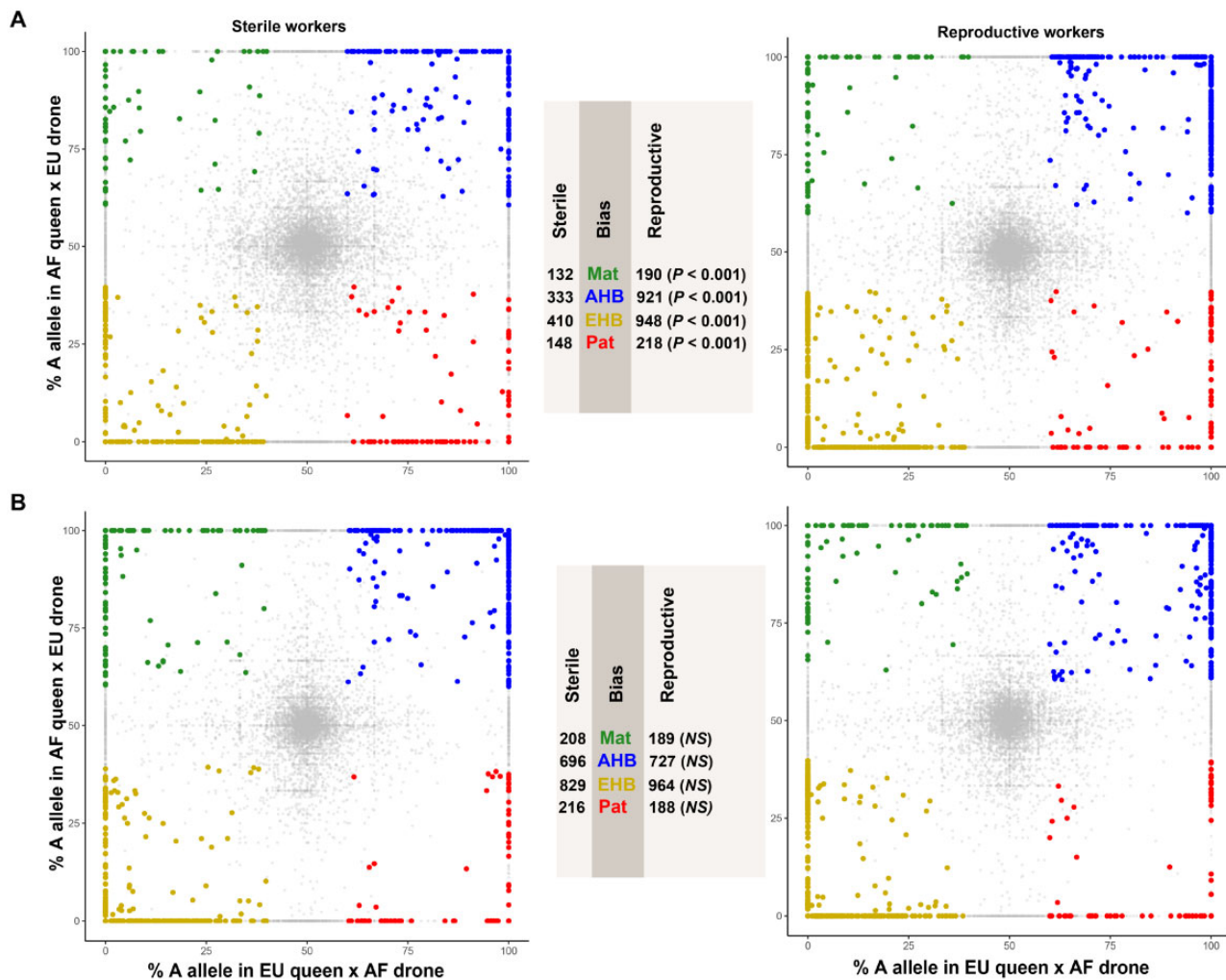


Fig. 2—DMPs for (A) workers in genetic block A and (B) workers in genetic block B represented by the relative percentage of Africanized methylation in each cross (Materials and Methods). Compared with sterile workers, reproductive workers in genetic block A showed an increased number of DMPs for each category allele-specific bias ($P < 0.05$ for all types). In contrast, this increase was not observed for any category of allele-specific bias in genetic block B. Each colored data point represents a specific type of allele-specific DMP—green is maternal, blue is Africanized, gold is European, red is paternal, and gray is not significant.

samples that is consistent with the effects of parent-of-origin and lineage (see Materials and Methods). The identified CpGs are referred to as DMPs and are summarized in table 1 based on their direction of bias. Figure 1 shows some examples of DMPs exhibiting lineage effects and parent-of-origin effects for both sterile and reproductive workers in each of the two blocks.

The strongest among the two factors was the lineage effect in both genetic blocks. In block A, a total of 743 and 1,869 individual CpG sites showed variation consistent with the lineage bias in sterile and reproductive workers, respectively. Similarly, block B had 1,525 and 1,691 individual sites showing lineage-based variation in sterile and reproductive workers, respectively (table 1). Interestingly, we observed greater bias toward European alleles—that is, there were a

greater number of European-biased CpGs than Africanized-biased CpGs. This trend was observed in both blocks and both sterile and reproductive workers, and statistically significant in all cases except the reproductive workers in block A (table 1; χ^2 test, $P < 0.05$).

We also found hundreds of DMPs in both reproductive states that showed parent-of-origin effects (table 1 and figs. 1A, B, E, and F and 2). For block A, we observed a total of 280 parent-of-origin DMPs in the sterile workers, of which 132 displayed maternal bias and 148 displayed paternal bias (table 1 and fig. 2). In the reproductive workers, we observed a total of 408 parent-of-origin DMPs, of which 190 DMPs were maternal biased whereas 218 DMPs were paternal biased, which was a significant increase in paternal-biased DMPs compared with maternal-biased DMPs (χ^2 test,

$P < 0.01$; [table 1](#) and [fig. 2](#)). For block B, there were 208 maternal-biased DMPs as opposed to 216 paternal-biased DMPs in the sterile workers, and 189 maternal and 188 paternal-biased DMPs in the reproductive workers ([table 1](#) and [fig. 2](#)). Interestingly, there were greater numbers of DMPs in reproductive workers compared with sterile workers for all categories in block A (X^2 test, $P < 0.05$ for all directions of bias), but there were no significant differences in the numbers of DMPs between reproductive and sterile workers for any of the categories block B.

We found significant overlaps of DMPs between workers of different reproductive states. For example, 69 parent-of-origin DMPs were shared between sterile and reproductive workers in block A, whereas 119 parent-of-origin DMPs were shared between sterile and reproductive workers in block B, both of which were highly significant enrichments compared with a null expectation of no association (Fisher’s exact test, $P < 0.01$ for both comparisons). Despite the significant overlaps, a large number of DMPs were specific to each reproductive state. In block A, 211 and 339 parent-of-origin DMPs were specific to sterile and reproductive workers, respectively. Similarly, there were 305 parent-of-origin DMPs specific to sterile workers and 258 parent-of-origin DMPs specific to reproductive workers. Furthermore, 189 sterile-specific and 191 reproductive-specific DMPs are shared across blocks, which is again significant compared with the expectation of no association (Fisher’s exact test, $P < 0.01$ for both comparisons). These results point to common, robust factors influencing genome-wide DNA methylation that are independent of the reproductive state and genetic block. Genes containing these DMPs are discussed in more detail in the next section.

Genes Harboring Parent-of-Origin-Mediated Differential Methylation Signatures

To infer functional consequences of differential DNA methylation, we defined DMGs as genes that contained DMPs exhibiting the same direction of allelic methylation bias (Materials and Methods). For example, parent-of-origin DMPs in block A were found across 179 and 230 genes in the sterile and

reproductive workers, respectively, and these genes are subsequently referred to as parent-of-origin DMGs ([table 2](#)). DMGs in other categories were similarly identified ([table 2](#)). The majority of parent-of-origin DMGs in both reproductive states and genetic blocks contained just a singular DMP (sterile average: 1.21 DMPs; reproductive average: 1.24 DMPs). The genes containing the most DMPs in the sterile workers, GB41581, were found in block B, though it does not have an annotation associated with it ([supplementary table 2, Supplementary Material](#) online). Block B contained the gene with the most DMPs in reproductive workers (GB44685), though this gene was also unannotated ([supplementary table 2, Supplementary Material](#) online). Interestingly, protein bicaudal D (GB42781) appears as a maternally biased DMG in both genetic blocks and was previously shown to exhibit both allele-specific methylation and expression in bumble bees (Lonsdale et al. 2017).

To take advantage of the information provided by the two different genetic blocks, we combined DMGs from both blocks for GO, pathway, and comparative analyses. GO terms enriched in different DMG categories are shown in [supplementary table 3, Supplementary Material](#) online. GO terms for sterile parent-of-origin DMGs included protein glycosylation, ATP binding functions, and involved in fatty acid degradation pathways ([supplementary table 3, Supplementary Material](#) online). Reproductive parent-of-origin DMGs were enriched for functions involving intracellular protein transport and mRNA surveillance pathways ([supplementary table 3, Supplementary Material](#) online).

We observed moderate but significant overlaps between parent-of-origin DMGs of the two reproductive states in both blocks. Thus, these were the genes which showed parent-of-origin effects in both sterile and reproductive workers. Specifically, there were 16 DMGs showing maternal bias (Fisher’s exact test, $P < 0.01$) and 30 DMGs showing paternal bias (Fisher’s exact test, $P < 0.01$) overlapping between sterile and reproductive workers in block A. In block B, there were 45 maternal DMGs (Fisher’s exact test, $P < 0.01$) and 35 paternal DMGs (Fisher’s exact test, $P < 0.01$) overlapping between sterile and reproductive workers. Though none of the overlapping gene sets was enriched for specific GO terms, they nevertheless mirrored the DMP results and reinforce the idea of a common set of genes that are differentially methylated due to parent-of-origin effects.

Interestingly, there was significant overlap between genes showing lineage differential methylation and parent-of-origin differential methylation ([fig. 3](#)). We found 46 DMGs exhibiting both lineage and parent-of-origin biases in block A sterile workers (Fisher’s exact test, $P < 0.01$) and 83 DMGs showing both biases in reproductive workers (Fisher’s exact test, $P < 0.01$; [fig. 3](#)). In block B, sterile workers and reproductive workers had 96 and 83 genes belonging to lineage and parent-of-origin DMGs. Functions of genes that show both types of allele-specific methylation did not deviate from the enriched

Table 2

DMGs Categorized by the Worker Reproductive State and Direction of Allele-Specific Bias

	Block A		Block B	
	Sterile	Reproductive	Sterile	Reproductive
Parent-of-origin				
Maternal bias	82	113	140	127
Paternal bias	97	117	126	106
Lineage				
Africanized bias	165	313	258	259
European bias	201	314	293	321

NOTE.—DMGs contain DMPs that show the same direction of allele-specific bias.

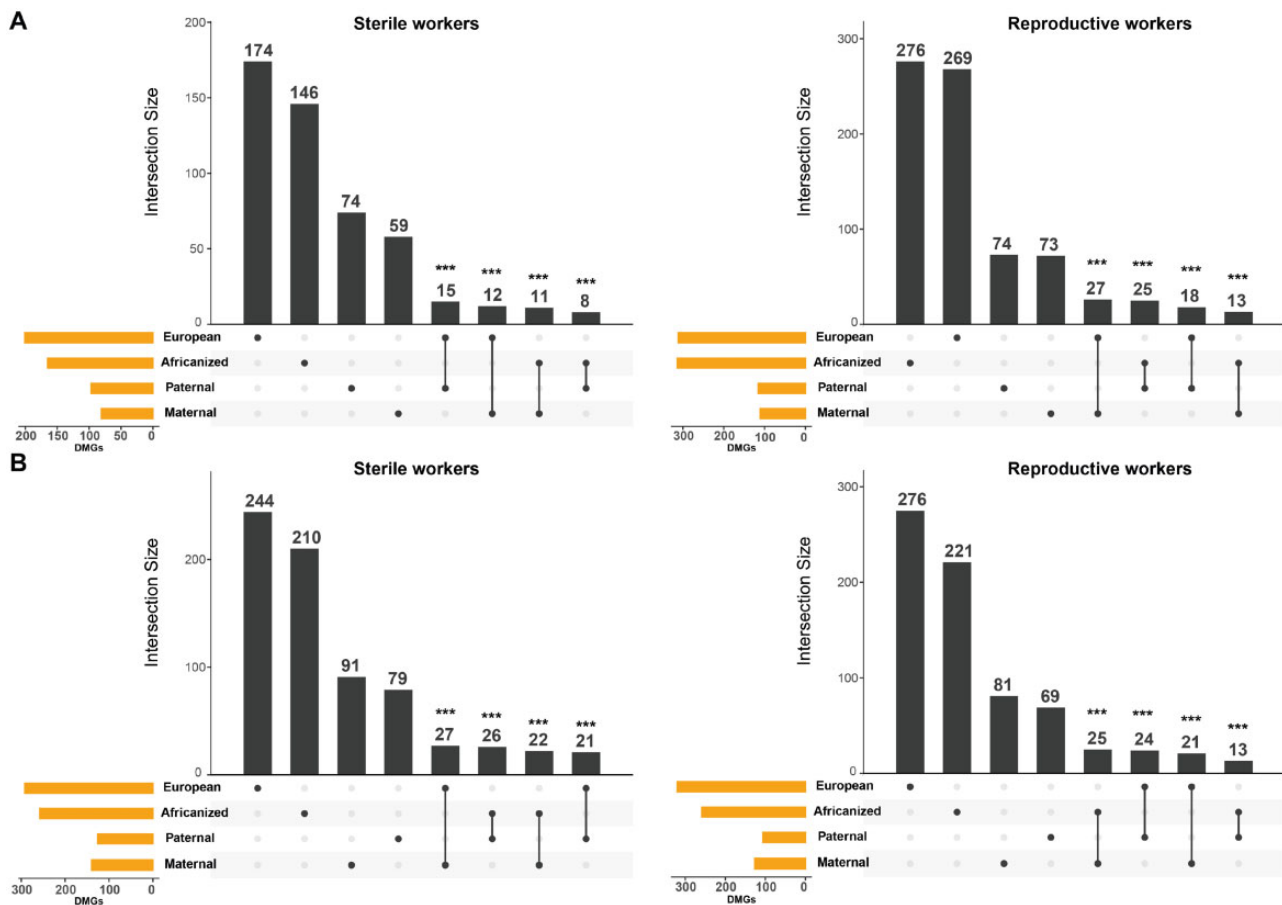


Fig. 3—The numbers of genes identified as a specific category of DMGs in sterile and reproductive workers in (A) genetic block A and (B) genetic block B. A number of genes were classified as both lineage and parent-of-origin DMGs. For example, in block A, 15 genes exhibited both European-biased and paternal-origin-biased methylation. ***Statistically significant overlap between groups at the $P < 0.001$ using a Fisher’s exact test.

GO terms of their respective reproductive states, which were generally focused on cell energy metabolism and signal transduction (supplementary table 3, Supplementary Material online). Because these genes exhibit both lineage and parent-of-origin differential methylation, they may be particularly labile in terms of allele-specific methylation.

We next examined parent-of-origin DMGs that were unique to sterile and reproductive workers to investigate the relationship between parent-specific methylation and reproductive phenotype. There were a total of 133 sterile-specific DMGs and 184 reproductive-specific DMGs in block A and 266 sterile-specific DMGs and 233 reproductive-specific DMGs in block B. Twelve such DMGs were commonly found in sterile workers of both blocks, whereas 22 DMGs were common between the reproductive workers (Fisher’s exact test, $P < 0.05$ for both comparisons). Although these overlaps were statistically significant, they did not exhibit any significant functional enrichment in our analysis, likely due to the small number. In comparison, DMGs specific to sterile workers in block A were enriched for GO terms associated with protein deubiquitination, whereas reproductive-worker-specific

DMGs were enriched for functions such as mRNA surveillance pathway and hydrolase activity (supplementary table 3, Supplementary Material online). For block B, sterile-specific DMGs were enriched for GO terms related to protein glycosylation and signal transduction, whereas reproductive-specific DMGs showed enriched GO terms such as intracellular transport (supplementary table 3, Supplementary Material online).

For additional validation of our results, we performed differential methylation analysis of gene body methylation by mirroring previous methods in insects (Lyko et al. 2010; Galbraith et al. 2015). These results largely recapitulated our findings from the DMP and DMG analyses, reproducing the same overall allele-specific methylation patterns and overlapping significantly with DMP containing genes (supplementary tables 6 and 7, Supplementary Material online).

Differential Allelic Methylation Is Not Strongly Associated with Differential Expression

To explore the link between parent-of-origin expression and methylation, we compared a previously obtained RNA-seq

data set (Galbraith, Kocher, et al. 2016) with our current results. The individuals from the RNA-seq study are sisters of the individuals in the current study. To make the results comparable to the methylation results, we reanalyzed the RNA-seq data using the same analysis pipeline as the current study (Materials and Methods). Our results recapitulated trends from the previous study, and although the number of genes in each category was different from the original study (possibly due to the different analysis methods used), they were all subsets of the genes from Galbraith, Kocher, et al. (2016) (supplementary table 4, Supplementary Material online). For both genetic blocks, there was a significantly more patrigene bias compared with matrigene bias as well as bias toward reproductive workers compared with sterile workers (Fisher's exact test, $P < 0.01$ for all comparisons). Interestingly, we found that DEGs varying due to parent-of-origin and lineage effects were almost exclusive to reproductive workers in both genetic blocks (supplementary table 4, Supplementary Material online). Additionally, there were essentially no overlap between allelic DMGs and allelic DEGs in either genetic blocks. In fact, the only overlap we observed was in reproductive workers for the lineage effect in block A and there was a complete lack of overlap for any parent-of-origin genes in both genetic blocks (supplementary table 5, Supplementary Material online).

Discussion

We used reciprocal crosses between divergent honey bee stocks and WGBS to understand factors that affect DNA methylation at genome-scale. We found a strong lineage effect on DNA methylation (tables 1 and 2). Previous studies indicated that genome-wide DNA methylation is highly affected by background genotypes in other species (Jones 2012; Smith and Meissner 2013; Mendizabal et al. 2014; Yi 2017; Keller et al. 2016), including the *Nasonia* jewel wasp (Wang et al. 2016) and bumble bee (Marshall et al. 2019). However, we also found that some of the CpGs in the honey bee genome show variation in methylation consistent with parent-of-origin effects, which was not observed in *Nasonia* (Wang et al. 2016). As far as we are aware, this is the first time a parent-of-origin effect has been demonstrated in an insect system, and it is unclear why they are present in honey bees and not in *Nasonia*: Clearly, additional studies in other insect species are needed. The numbers of DMPs and DMGs showing a parent-of-origin effect were 2–3-fold smaller than those exhibiting lineage effects, indicating that parent-of-origin effect is not as strong as genetic background effects. Nevertheless, the numbers of genes exhibiting parent-of-origin effects range between 3.2% and 9.9% of genes analyzed, similar ranges as observed in mammals (Luedi et al. 2007; Ferguson-Smith and Bourc'his 2018). We also observed that many genes harbored both parent-of-origin DMPs and lineage-specific DMPs in both blocks (fig. 3). This observation

could potentially indicate that some positions or some genes in the honey bee genome tend to be labile in terms of epigenetic modification, and potentially targets of regulation for a many different factors. Future studies using targeted deep sequencing of candidate loci in a broader range of genotypes and samples could more clearly define the methylation patterns at these sites.

Interestingly, we found that, with the exception of the paternal category, there was an increase in both DMP and DMG numbers in the reproductive workers compared with the sterile workers (X^2 test, $P < 0.05$ for all comparisons) in block A. This observation mirrored the increase of parent-of-origin effect in reproductive workers at the level of gene expression (Galbraith, Kocher, et al. 2016). However, in block B, this pattern was not observed (except a modest increase in European-biased DMPs, table 1, X^2 test, $P < 0.05$). One possibility is that this difference could have arisen due to the different ages of the workers between the two genetic blocks—though all the reproductive workers were confirmed to have activated ovaries, because workers in block A were 4 days older, they were likely more reproductively mature, which could manifest in clearer DNA methylation difference between worker castes. However, this hypothesis needs to be reexamined with larger data set that spans a broader range of ages in the future. Given previous evidence of tissue-specific methylation imprints (Weinstein 2001; Babak et al. 2015), another possibility may have been the signal from the methylome of the eggs within developed ovarioles which were not removed prior to DNA extraction and have been shown to have significantly different methylation profiles than adult honey bees (Drewell et al. 2014).

Previous work on parent-of-origin gene expression supported the prediction that worker ovary activation was associated with biased expression of patrigenes, with a stronger paternal bias in reproductive workers compared with sterile workers (Galbraith et al. 2015). Our reanalysis of the RNA-seq data recapitulated this finding, though we did not see the same patterns in our DNA methylation analysis. In terms of the link between DNA methylation and gene expression, we observed almost no overlap between parent-specific gene expression and methylation. This could indicate that either DNA methylation does not affect parent-of-origin gene expression or the effect of DNA methylation is indirect. Several genes we found exhibiting allele-specific methylation were related to chromatin structure and remodeling (histone-lysine *N*-methyltransferases, chromatin-remodeling complex ATPase chain *lswi*; supplementary table 2, Supplementary Material online), which may affect changes in chromatin structure that can affect gene expression levels without methylation changes to the gene (Lachner et al. 2001; Prantero and Bongiorno 2012). For example, histone-lysine methylation was shown to be implicated in chromosome silencing in two insect species (Prantero and Bongiorno 2012). It is worth noting that studies in insects thus far suggest that differential DNA

methylation does not directly correlate with differential gene expression (Galbraith et al. 2015; Arsenault et al. 2018; Wu et al. 2020). Rather, DNA methylation may affect other aspects of gene expression such as gene expression variability or alternative splicing (Huh et al. 2013; Hunt et al. 2013; Wang et al. 2013; Galbraith et al. 2015; Arsenault et al. 2018). Although our current data set is insufficient for distinguishing these alternative scenarios, our results call for future studies to examine other epigenetic marks, particularly histone modifications and/or open chromatin features in experiments utilizing greater number of allele-specific SNPs. Such studies can expand the number of sites identified that exhibit allele-specific methylation and will aid in understanding the general principle of genomic imprinting as well as the main role of DNA methylation in insects.

Acknowledgments

We would like to thank Tom Glenn (retired, Glenn Apiaries) for generating the honey bee samples used in this study. This work was supported by the National Science Foundation (Grant Nos. MCB-1616257 to C.M.G. and MCB-1615664 to S.V.Y.). The authors declare no competing interests.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Literature Cited

- Anders S, Pyl PT, Huber W. 2015. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31(2):166–169.
- Arsenault SV, Hunt BG, Rehan SM. 2018. The effect of maternal care on gene expression and DNA methylation in a subsocial bee. *Nat Commun.* 9(1):3468.
- Babak T, et al. 2015. Genetic conflict reflected in tissue-specific maps of genomic imprinting in human and mouse. *Nat Genet.* 47(5):544–549.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol.* 57(1):289–300.
- Bird A. 2002. DNA methylation patterns and epigenetic memory. *Genes Dev.* 16(1):6–21.
- Bonasio R, et al. 2012. Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Curr Biol.* 22(19):1755–1764.
- Drewell RA, et al. 2014. The dynamic DNA methylation cycle from egg to sperm in the honey bee *Apis mellifera*. *Development* 141(13):2702–2711.
- Elango N, Hunt BG, Goodisman MA, Yi SV. 2009. DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. *Proc Natl Acad Sci U S A.* 106(27):11206–11211.
- Elsik CG, et al. 2014. Finding the missing honey bee genes: lessons learned from a genome upgrade. *BMC Genomics.* 15(1):86.
- Ferguson-Smith AC. 2011. Genomic imprinting: the emergence of an epigenetic paradigm. *Nat Rev Genet.* 12(8):565–575.
- Ferguson-Smith AC, Bourc'his D. 2018. The discovery and importance of genomic imprinting. *Elife* 7:e42368.
- Galbraith DA, Kocher SD, et al. 2016. Testing the kinship theory of intragenomic conflict in honey bees (*Apis mellifera*). *Proc Natl Acad Sci U S A.* 113(4):1020–1025.
- Galbraith DA, Yang X, Niño EL, Yi S, Grozinger C. 2015. Parallel epigenomic and transcriptomic responses to viral infection in honey bees (*Apis mellifera*). *PLoS Pathog.* 11(3):e1004713.
- Galbraith DA, Yi SV, Grozinger CM. 2016. Evaluation of possible proximate mechanisms underlying the kinship theory of intragenomic conflict in social insects. *Integr Comp Biol.* 56(6):1206–1214.
- Gibson JD, Arechavala-Velasco ME, Tsuruda JM, Hunt GJ. 2015. Biased allele expression and aggression in hybrid honeybees may be influenced by inappropriate nuclear-cytoplasmic signaling. *Front Genet.* 6:343.
- Haig D. 1992. Intragenomic conflict and the evolution of eusociality. *J Theor Biol.* 156(3):401–403.
- Haig D. 2000. The kinship theory of genomic imprinting. *Annu Rev Ecol Syst.* 31(1):9–32.
- Huang da W, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 4(1):44–57.
- Huh I, Wu X, Park T, Yi SV. 2019. Detecting differential DNA methylation from sequencing of bisulfite converted DNA of diverse species. *Brief Bioinf.* 20(1):33–46.
- Huh I, Zeng J, Park T, Yi SV. 2013. DNA methylation and transcriptional noise. *Epigenet Chromatin* 6(1):9.
- Hunt BG, Glastad KM, Yi SV, Goodisman MA. 2013. Patterning and regulatory associations of DNA methylation are mirrored by histone modifications in insects. *Genome Biol Evol.* 5(3):591–598.
- Jones PA. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet.* 13(7):484–492.
- Keller TE, Lasky JR, Yi SV. 2016. The multivariate association between genomewide DNA methylation and climate across the range of *Arabidopsis thaliana*. *Mol Ecol.* 25(8):1823–1837.
- Kocher SD, et al. 2015. A search for parent-of-origin effects on honey bee gene expression. *G3 (Bethesda)* 5:1657–1662.
- Krueger F, Andrews SR. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27(11):1571–1572.
- Krueger F, Andrews SR. 2016. SNPsplit: allele-specific splitting of alignments between genomes with known SNP genotypes. *F1000Research* 5:1479.
- Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T. 2001. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* 410(6824):116–120.
- Law JA, Jacobsen SE. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet.* 11(3):204–220.
- Lindsey ARI, et al. 2018. Comparative genomics of the miniature wasp and pest control agent *Trichogramma pretiosum*. *BMC Biol.* 16(1):54.
- Lonsdale Z, et al. 2017. Allele specific expression and methylation in the bumblebee, *Bombus terrestris*. *PeerJ* 5:e3798.
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15(12):550.
- Luedi PP, et al. 2007. Computational and experimental identification of novel human imprinted genes. *Genome Res.* 17(12):1723–1730.
- Lyko F, et al. 2010. The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS Biol.* 8(11):e1000506.
- Marshall H, Lonsdale ZN, Mallon EB. 2019. Methylation and gene expression differences between reproductive and sterile bumblebee workers. *Evol Lett.* 3(5):485–499.
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17(1):10–12.
- Mendizabal I, Keller TE, Zeng J, Yi SV. 2014. Epigenetics and evolution. *Integr Comp Biol.* 54(1):31–42.

- Oldroyd BP, et al. 2014. A parent-of-origin effect on honeybee worker ovary size. *Proc Biol Sci.* 281(1775):20132388.
- Park Y, Wu H. 2016. Differential methylation analysis for BS-seq data under general experimental design. *Bioinformatics* 32(10):1446–1453.
- Patten MM, et al. 2014. The evolution of genomic imprinting: theories, predictions and empirical tests. *Heredity (Edinb).* 113(2):119–128.
- Pegoraro M, Marshall H, Lonsdale ZN, Mallon EB. 2017. Do social insects support Haig's kin theory for the evolution of genomic imprinting? *Epigenetics* 12(9):725–742.
- Prantera G, Bongiorno S. 2012. Mealybug chromosome cycle as a paradigm of epigenetics. *Genet Res Int.* 2012:1–11.
- Queller DC. 2003. Theory of genomic imprinting conflict in social insects. *BMC Evol Biol.* 3(1):15.
- Reik W, Walter J. 2001. Genomic imprinting: parental influence on the genome. *Nat Rev Genet.* 2(1):21–32.
- Schultz MD, Schmitz RJ, Ecker JR. 2012. 'Leveling' the playing field for analyses of single-base resolution DNA methylomes. *Trends Genet.* 28(12):583–585.
- Smith ZD, Meissner A. 2013. DNA methylation: roles in mammalian development. *Nat Rev Genet.* 14(3):204–220.
- Urich MA, Nery JR, Lister R, Schmitz RJ, Ecker JR. 2015. MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. *Nat Protoc.* 10(3):475–483.
- Wang X, Werren JH, Clark AG. 2016. Allele-specific transcriptome and methylome analysis reveals stable inheritance and cis-regulation of DNA methylation in *Nasonia*. *PLoS Biol.* 14(7):e1002500.
- Wang X, et al. 2013. Function and evolution of DNA methylation in *Nasonia vitripennis*. *PLoS Genet.* 9(10):e1003872.
- Wang Y, et al. 2006. Functional CpG methylation system in a social insect. *Science* 314(5799):645–647.
- Weinstein LS. 2001. The role of tissue-specific imprinting as a source of phenotypic heterogeneity in human disease. *Biol Psychiatry* 50(12):927–931.
- Wilkins JF, Haig D. 2003. What good is genomic imprinting: the function of parent-specific gene expression. *Nat Rev Genet.* 4(5):359–368.
- Wolf JB, Hager R. 2006. A maternal-offspring coadaptation theory for the evolution of genomic imprinting. *PLoS Biol.* 4(12):e380.
- Wu X, et al. 2020. Distinct epigenomic and transcriptomic modifications associated with *Wolbachia*-mediated asexuality. *PLoS Pathog.* 16(3):e1008397.
- Yi SV. 2017. Insights into epigenome evolution from animal and plant methylomes. *Genome Biol Evol.* 9(11):3189–3201.
- Zeng J, Yi SV. 2010. DNA methylation and genome evolution in honeybee: gene length, expression, functional enrichment covary with the evolutionary signature of DNA methylation. *Genome Biol Evol.* 2:770–780.

Associate editor: Rachel O'Neill