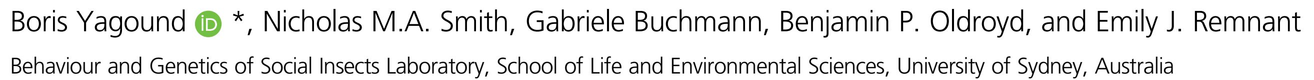
Unique DNA Methylation Profiles Are Associated with *cis*-Variation in Honey Bees


Boris Yagound *, Nicholas M.A. Smith, Gabriele Buchmann, Benjamin P. Oldroyd, and Emily J. Remnant

Behaviour and Genetics of Social Insects Laboratory, School of Life and Environmental Sciences, University of Sydney, Australia

*Corresponding author: E-mail: boris.yagound@sydney.edu.au.

Accepted: August 9, 2019

Data deposition: This project has been deposited at the NCBI Sequence Read Archive under the accession PRJNA489158.

Abstract

DNA methylation is an important epigenetic modification that mediates diverse processes such as cellular differentiation, phenotypic plasticity, and genomic imprinting. Mounting evidence suggests that local DNA sequence variation can be associated with particular DNA methylation states, indicating that the interplay between genetic and epigenetic factors may contribute synergistically to the phenotypic complexity of organisms. Social insects such as ants, bees, and wasps have extensive phenotypic plasticity manifested in their different castes, and this plasticity has been associated with variation in DNA methylation. Yet, the influence of genetic variation on DNA methylation state remains mostly unknown. Here we examine the importance of sequence-specific methylation at the genome-wide level, using whole-genome bisulfite sequencing of the semen of individual honey bee males. We find that individual males harbor unique DNA methylation patterns in their semen, and that genes that are more variable at the epigenetic level are also more likely to be variable at the genetic level. DNA sequence variation can affect DNA methylation by modifying CG sites directly, but can also be associated with local variation in *cis* that is not CG-site specific. We show that covariation in sequence polymorphism and DNA methylation state contributes to the individual-specificity of epigenetic marks in social insects, which likely promotes their retention across generations, and their capacity to influence evolutionary adaptation.

Key words: DNA methylation, epigenetics, *Apis mellifera*, allele-specific methylation, phenotypic plasticity.

Introduction

Complex interactions between genetic and epigenetic factors shape the diversity of organismal phenotypes, from cell differentiation in multicellular eukaryotes to caste determination in insect societies (West-Eberhard 2003). Epigenetic mechanisms act as a major source of phenotypic plasticity by triggering changes in gene function without altering the DNA sequence (Berger et al. 2009). Epigenetic modifications are found across plants, animals, and fungi, and play major roles in processes as diverse as development, behavior, and disease (Goldberg et al. 2007; Bonasio et al. 2010). They can modify genome function through an individual's ontogeny and can be transmitted across generations (Richards 2006; Heard and Martienssen 2014; Miska and Ferguson-Smith 2016).

One of the most important epigenetic mechanisms involves the methylation of genomic DNA. DNA methylation is a covalent modification that occurs by the addition of a methyl group to the fifth carbon of cytosines, mostly in CG dinucleotides (Klose and Bird 2006; Suzuki and Bird 2008; Zemach et al. 2010), although adenine methylation can also occur (Ratel et al. 2006). DNA methylation is controlled by DNA methyltransferases (DNMTs). DNMT1 is responsible for maintaining methylation states across cell divisions, whereas DNMT3 is involved in de novo methylation (Klose and Bird 2006), although these functions can overlap (Jeltsch and Jurkowska 2014). DNA methylation can be reversible, in particular through the action of the ten–eleven translocation (TET) family enzymes (Kohli and Zhang 2013). It is largely unclear how de novo DNMTs interact with cofactors to identify specific targets in the genome, however the process may be linked to chromatin state and other epigenetic factors like histone modifications and transcription factor binding (Do et al. 2017).

Recent studies in mammals have shown that underlying genotype can have a strong influence on DNA methylation state throughout the genome (Schilling et al. 2009; Meaburn et al. 2010; Shoemaker et al. 2010; Tycko 2010; Lienert et al. 2011; Fang et al. 2012). Allele-specific methylation, where

© The Author(s) 2019. 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

methylation state depends on the local DNA sequence in *cis*, is becoming increasingly recognized as playing an important role in genome function (Tycko 2010; Do et al. 2017). *Cis*-mediated allele-specific methylation is distinct from other forms of epigenetic allelic asymmetries, such as genomic imprinting (Do et al. 2017). Genomic imprinting corresponds to differential expression of alleles in a parent-of-origin manner (Reik and Walter 2001). Imprinting probably originates from evolutionary conflicts between paternally-derived and maternally-derived genes regarding their effects on kin (Haig 2000). Genomic imprinting is often mediated by DNA methylation in mammals and plants, where alleles are differentially methylated based on the parent-of-origin (Tycko 2010; Suzuki and Bird 2008). In humans, *cis*-mediated allele-specific methylation is prevalent across genes and tissues, but only a small minority of genes are imprinted (Meaburn et al. 2010; Tycko 2010). Both forms of allele-specific methylation have important and distinct consequences in the context of epigenetic inheritance (Kerkel et al. 2008; Tycko 2010; Heard and Martienssen 2014; Mendizabal et al. 2014). Genomic imprinting is non-Mendelian (i.e., methylation state depends on the parent-of-origin from which the allele comes), whereas *cis*-mediated allele-specific methylation follows Mendelian inheritance patterns (i.e., methylation state depends on the genotype of the allele in *cis*). The interplay between genetic and epigenetic factors could thus be key to understanding the emergence of phenotypic plasticity and the inheritance of epigenetic marks (Kerkel et al. 2008; Tycko 2010; Heard and Martienssen 2014; Mendizabal et al. 2014).

Social insects (ants, termites, some bees, and wasps) are particularly relevant for investigating the evolutionary implications of genetic and epigenetic interactions (Li-Byarlay et al. 2013; Bonasio 2014; Mendizabal et al. 2014; Welch and Lister 2014; Glastad et al. 2015, 2019; Yan et al. 2015; Maleszka 2016; Cridge et al. 2017). DNA methylation in social insects is sparse and primarily intragenic, occurring mostly in exons (Suzuki and Bird 2008; Lyko et al. 2010; Zemach et al. 2010; Bonasio et al. 2012; Smith et al. 2012; Terrapon et al. 2014). Despite being found at low levels genome-wide, DNA methylation affects many genes and is thought to play important roles in phenotypic plasticity and epigenetic inheritance. Mechanistically, DNA methylation has been suggested to regulate gene function via alternative splicing (Li-Byarlay et al. 2013; but see Arsenault et al. 2018). Social insects display extreme intraspecific variability in morphology, behavior, and physiology (Oster and Wilson 1978; Simpson et al. 2011). This phenotypic plasticity is under epigenetic control, as all individuals typically originate from the same genome (Schwander et al. 2010). In social insects, DNA methylation has been linked with phenotypic plasticity, particularly in the context of caste determination (Glastad et al. 2011, 2014; Herb et al. 2012; Weiner and Toth 2012; Yan et al. 2015; Li-Byarlay 2016).

DNA methylation has also been suggested to be a potential epigenetic mediator in the context of genomic imprinting (Queller 2003; Drewell et al. 2012). In social insects where the queen mates multiply (i.e., polyandry), males can increase their fitness if their own daughters reproduce. This is predicted to lead to genomic imprinting, whereby fathers differentially "imprint" their genomes in order to make their daughter offspring more likely to reproduce. In contrast, the queen is equally related to all the workers and should favor her own reproduction. Theory thus predicts conflicts of interest between paternally-derived and maternally-derived genes affecting worker reproduction (Queller 2003; Drewell et al. 2012). Parent-of-origin effects have been observed in worker honey bees regarding ovary size (Oldroyd et al. 2014; Reid et al. 2017) and gene expression (Kocher et al. 2015; Galbraith et al. 2016), supporting the existence of genomic imprinting in social insects. Yet the underlying proximate mechanisms remain undetermined.

Beyond imprinting, the existence of polyandry creates a number of worker patrilines in each colony, each sired by a different father (Palmer and Oldroyd 2000). Workers from distinct patrilines typically perform different tasks (Guzman-Novoa et al. 2005; Oldroyd and Fewell 2007) and differ widely in their reproductive traits (Yagound et al. 2017). The potential for males to transfer epigenetic information to their progeny could potentially lead to patriline-dependent worker traits, particularly if these epigenetic factors act synergistically with genetic variation (Smith et al. 2008).

The interplay between genotype and methylation state is rarely considered in social insect methylation studies (Maleszka 2016; Remnant et al. 2016), despite its prevalence in other taxa (see above). Allele-specific methylation is known to occur in insects (Wang et al. 2016), particularly honey bees (Remnant et al. 2016; Wedd et al. 2016). Further, sample-specific differences in DNA methylation patterns can be of similar magnitude to caste-specific differences (Libbrecht et al. 2016). Covariation between genetic and epigenetic factors could thus play a key role in determining the specificity of epigenetic marks.

In this study, we determine the whole-genome methylation profile of the semen of seven unrelated honey bee males, and investigate the pervasiveness of sequence-specific methylation. Semen is particularly relevant in the context of epigenetic inheritance (Rando 2016; Tillo et al. 2016) and has a methylome that is distinct from that of eggs and adult males (Drewell et al. 2014). Honey bee males are haploid, so all sperm produced by any individual is genetically identical. Thus, the use of semen from individual haploid males allowed us to determine methylation state in a homogeneous genetic background.

Materials and Methods

Experimental Design

This study comprises three bisulfite sequencing experiments. In the first experiment, we determined the whole-genome methylation profile of the semen of three honey bee males using a protocol that yielded a 5–6-fold average genome coverage. We refer to these samples as low coverage (LC) samples. In the second experiment, we determined the whole-genome methylation profile of the semen of an additional four honey bee males using a modified protocol that yielded a 30–31-fold coverage. We refer to these samples as high coverage (HC) samples. The HC samples served to confirm and expand the LC findings, but also allowed us to clarify whether sequencing depth has a significant effect on our ability to accurately quantify methylation. In the third experiment, we performed additional validation by bisulfite PCR sequencing for six genes showing differential methylation and sequence variation in silico in the first two experiments.

Whole-Genome Bisulfite Sequencing

For LC samples, three sexually-mature male honey bees were sampled from three different colonies of standard Australian commercial stock of *Apis mellifera* (mainly *A. m. ligustica*) maintained at the University of Sydney, Australia, in December 2016. For HC samples, four sexually-mature male honey bees were sampled from four different colonies of *A. m. capensis* located in Stellenbosch, South Africa, in November 2016. Honey bee males complete spermatogenesis during the pupal stage (Koeniger et al. 2014). Mature sperm is then stored by adult males until the opportunity to mate arises, after which they die. Semen was collected from each male's ejaculate following eversion of the endophallus, and immediately stored at −80 °C for sequencing. The use of individual semen samples allowed us to eliminate genetic heterogeneity within each sample, while ensuring that each sample was an independent replicate (Libbrecht et al. 2016). Sequencing samples from different honey bee subspecies also allowed us to verify that our conclusions are likely of general significance.

For LC samples, DNA was isolated using standard phenol/chloroform/isoamyl alcohol extraction protocols (Sambrook et al. 1989). We performed bisulfite conversion and library preparation using 100 ng of input DNA with the Pico Methyl-Seq Library Prep Kit (Zymo Research) according to the manufacturer's instructions. Bisulfite treatment converts non-methylated cytosines into thymines, whereas methylated cytosines are protected from conversion. Unmethylated lambda phage DNA (0.1% w/w, Promega) was used as a spike-in control to assess conversion efficiency. Libraries were quantified with Qubit fluorometer (Life Technologies). Libraries were sequenced at the Australian Genome Research Facility (Melbourne) on an Illumina HiSeq2500 (one lane of 50 bp single-end sequencing).

For HC samples, DNA isolation, quantification, and lambda phage DNA spike-in was performed as described above. We performed library preparation using NEBNext Ultra II DNA Library Prep Kit for Illumina (New England BioLabs) according to the manufacturer's instructions. Briefly, DNA was sheared to 250 bp target fragments on a Covaris E220 sonicator with the following settings: 20% duty factor, 50 cycles/bursts, 18 W peak incident power, 80 s duration. DNA fragments were end-repaired, A-tailed and ligated with NEBNext methylated adapter oligos for Illumina (New England BioLabs). We performed bisulfite conversion using EZ DNA Methylation-Direct Kit (Zymo Research) following the manufacturer's instructions. Converted DNA fragments were amplified for seven cycles with KAPA HiFi HotStart Uracil+ Readymix (Roche). After a final cleanup with Agencourt AMPure XP beads (Beckman Coulter), we checked the size distribution and molarity of the library on a TapeStation (Agilent Technologies), and quantified the library concentration on a Qubit fluorometer (Life Technologies). Libraries were sequenced at the Australian Genome Research Facility (Melbourne) on HiSeq2500 system (Illumina) using one lane of 125 bp paired-end sequencing. WGBS data for the three LC semen samples (i.e., Sperm i3, Sperm i4, and Sperm i7) and the four HC semen samples (i.e., Sperm i47, Sperm i49, Sperm i51, and Sperm i53) have been deposited to the NCBI Sequence Read Archive under accession number PRJNA489158.

Data Processing

We checked the quality of the raw data with FastQC 0.11.15 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc; last accessed February 27, 2019). Low quality reads and adapter sequences were trimmed using TrimGalore 0.4.1 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore; last accessed September 4, 2018). For LC samples, we used default parameters, removing low quality reads (Phred score <20) and short reads (length <20 bp) with an additional trimming of the first 6 bp to control for GC content bias. For HC samples, we used more stringent parameters, removing low quality reads (Phred score <30) and short reads (length <36 bp) with an additional trimming of the first 3 bp to control for GC content bias. Remaining reads were mapped to the honey bee reference genome assembly Amel_4.5 (The Honeybee Genome Sequencing Consortium 2006) using Bismark 0.16.1 (Krueger and Andrews 2011) with Bowtie2 2.2.9 (Langmead and Salzberg 2012), using the "nondirectional" option for LC samples only. Duplicated reads were then removed with Bismark and methylation-biased reads were determined and removed using BWASP (https://github.com/BrendelGroup/BWASP; last accessed December 15, 2018). See supplementary tables

S1 and S2, Supplementary Material online, for sequencing statistics on each library.

Methylation Analysis

Methylation calling was performed with Bismark. For LC samples, any cytosine with at least four reads was regarded as sufficiently covered. This threshold was chosen following BWASP which determines the minimal coverage necessary for a site to be significantly methylated. For HC samples we considered CG sites with at least ten reads to be sufficiently covered. Methylation level at each site was calculated as the proportion of C to (C + T) reads (Glastad et al. 2017). Significantly methylated sites were then calculated using a binomial probability model, using the bisulfite conversion rate for each sample as the probability of success, and performing Bonferroni corrections at the 1% significance level using BWASP (Standage et al. 2016). For HC samples we excluded mCGs with a coverage >500×. We focused on CG methylation as it is the most prevalent context of genomic methylation (Klose and Bird 2006; Suzuki and Bird 2008; Zemach et al. 2010), although our samples also contained non-CG methylated sites (supplementary tables S3 and S4, Supplementary Material online). Genomic features were annotated with HOMER 4.9.1 (Heinz et al. 2010) using the *A. mellifera* official gene set amel_OGSv3.2 (Elsik et al. 2014). Methylation level of each genomic and genic feature was calculated as the average methylation level (see above) for all sufficiently covered sites across that particular feature (Glastad et al. 2017). Methylation level of each gene was calculated as the average methylation level for all sufficiently covered sites across that particular gene. Methylated genes were defined as any gene containing at least four mCGs and having a methylation level ≥5%, as determined by plotting genic methylation levels for each sample and selecting an appropriate cutoff threshold (supplementary fig. S1, Supplementary Material online; Arsenault et al. 2018).

Gene Ontology analyses were performed using DAVID (Huang da et al. 2009). *P* values were corrected for multiple comparisons following the Benjamini and Hochberg procedure (Benjamini and Hochberg 1995).

We determined whether an mCG found in one sample was also methylated at the same site in another sample using custom scripts in R 3.3.3 (R Core Team 2017). We required that the site had sufficient coverage in both samples to ensure that the observed differences in methylation frequency were not an artifact of coverage differences. Differentially methylated cytosines (DMCs) were calculated for each possible pairwise comparison among LC and HC samples using the R package methylKit 1.0.0 (Akalin et al. 2012). CG sites were considered differentially methylated if their methylation difference was 25% or greater between the two samples, and their *q*-value (Fisher's exact test corrected *P* value; Wang et al. 2011) was 0.01 or less (Remnant et al. 2016).

Differentially methylated genes (DMGs) were defined as any gene containing at least one DMC for LC samples, and at least ten DMCs for HC samples (upscaling for HC samples to account for the order of magnitude difference in coverage). For each individual semen sample, a list of DMGs were compiled containing all genes showing differential methylation with any of the other two (LC) or three (HC) samples. Consistently methylated genes (CMGs) were defined as any methylated gene that did not contain any DMCs. Methylation difference for each DMG of each pairwise comparison was calculated as the average methylation across all DMCs for one sample subtracted by the average methylation across all DMCs for the other sample. To assess whether certain genes frequently came up as differentially methylated, we calculated the proportion of DMGs that were consistently present in each pairwise comparison from the total number of unique DMGs found when considering all pairwise comparisons together.

SNP Analysis

Single nucleotide polymorphisms (SNPs) were called using BISCUIT 0.2.2 (https://github.com/zwdzwd/biscuit; last accessed October 11, 2018) which allows calling of variants from bisulfite sequencing data. SNPs were filtered with VCFtools 0.1.15 (Danecek et al. 2011). Hard trimming for LC samples consisted of removing variants which had a read depth (DP) <4, an allele support (SP) <4, had ambiguous alternative alleles (i.e., neither A nor G), were heterozygous (unexpected in haploid samples), and were found within nonnuclear regions. For HC samples, we excluded variants that had a quality score (QUAL) <30, a genotype quality (GQ) <20, a read depth (DP) <10 and >500, an allele support (SP) <10, had ambiguous alternative alleles, were heterozygous, and were found within nonnuclear regions. We then determined *cis*-SNPs as any SNP that was found within 1 kb upstream or downstream of an mCG.

Direct Bisulfite PCR Validation

We collected eight male honey bees at a natural mating lek (drone congregation area; Koeniger et al. 2014) located at the University of Sydney using a Williams drone trap (Williams 1987). Briefly, males were lured inside a net containing baits impregnated with artificial queen pheromone (E)-9-oxodec-2-enoic acid (9-ODA) and suspended from a helium balloon. Semen was collected as described above. DNA was extracted from each semen sample using the DNeasy blood and tissue kit (Qiagen), and DNA was bisulfite converted using the EZ DNA Methylation-Direct kit (Zymo Research), according to the manufacturer's instructions.

We selected six genes with sequence-specific methylation states in the WGBS experiments to perform bisulfite PCR sequencing, to test the hypothesis that particular sequence polymorphisms are associated with particular methylation states: *oxysterol-binding protein-related protein 2* (GB52517), *protocadherin-like wing polarity protein stan* (GB51276), *atrial natriuretic peptide-converting enzyme* (GB54775), *Glutamate receptor ionotropic, NMDA 3A* (GB41839), *heat shock protein Hsp70Ab-like* (GB50609); and *parafibromin* (GB43888). Nested primers were designed spanning regions containing differentially methylated sites within each gene (800–1000 bp; supplementary table S5, Supplementary Material online). PCR products were amplified using two-step nested PCR and KAPA 2G Robust DNA polymerase (KAPA Biosystems) (Remnant et al. 2016). We included T7 and T3 tags at the 5′-ends of the second nested set of primers to facilitate direct sequencing from PCR products. PCR products were sequenced by Macrogen (South Korea). Sequences were quality trimmed and manually examined for methylation using Geneious 10.2.4 (Kearse et al. 2012).

We examined each gene for sequence polymorphisms by sequencing PCR products of nonbisulfite converted DNA. T7- and T3-tagged primers were designed to amplify 1–2 kb regions of each gene that extended beyond the amplified bisulfite-PCR fragments (supplementary table S5, Supplementary Material online). This enabled sorting of each individual semen sample into distinct alleles based on SNP and insertion/deletion variation, particularly for alleles where variable sites were masked by bisulfite treatment or occurred outside of the bisulfite-PCR fragment.

Methylation density for each fragment was calculated as the proportion of mCGs relative to the total number of CG sites. To determine whether individuals with the same alleles had similar methylation patterns we performed agglomerative hierarchical clustering of methylation patterns (with each CG site characterized as being either methylated or unmethylated) in all eight individuals using the R package pvclust 2.0.0 (Suzuki and Shimodaira 2006). This analysis was based on correlation distances with the complete agglomerative method following Arsenault et al. (2018), and was restricted to genes containing a maximum of three alleles (i.e., *protocadherin-like wing polarity protein stan*, *glutamate receptor ionotropic, NMDA 3A, atrial natriuretic peptide-converting enzyme*; and *oxysterol-binding protein-related protein 2*).

Results

Whole-Genome Bisulfite Sequencing

Bisulfite sequencing libraries yielded a total of 190.5 million reads (9.52 Gb), with 5.2–5.9-fold genome coverage for the three LC samples (Sperm i3, i4, and i7), and a total of 187.5 million reads (47.26 Gb), with 29.9–31.0-fold genome coverage for the four HC samples (Sperm i47, i49, i51, and i53) (supplementary tables S1 and S2, Supplementary Material online). Approximately 10–11 million cytosines in a CG context were sufficiently covered for analysis in all samples (table 1 and supplementary table S6, Supplementary Material online). Of these, the number of significantly methylated CGs (mCGs) in LC samples was 16,921–21,179, giving genome-wide methylation densities of ∼0.2% (supplementary table S6, Supplementary Material online). In HC samples, there were 122,423–127,050 mCGs, giving genome-wide methylation densities of ∼1.1% (table 1).

As previously established in honey bees (Lyko et al. 2010; Zemach et al. 2010; Drewell et al. 2014), methylation was sparse across the genome (supplementary figs. S2 and S3, Supplementary Material online), and the vast majority of mCGs were found within genes, particularly in exons, which accounted for over two-thirds of mCGs (supplementary table S7, Supplementary Material online). In the HC samples, methylation levels were significantly higher in exons compared with other genic and intergenic regions (Kruskal–Wallis rank sum tests with Dunn's post hoc tests: all $P < 0.00001$; supplementary fig. S4, Supplementary Material online), and intragenic methylation levels were always higher in exons than in introns (all $P < 0.00001$; supplementary fig. S5, Supplementary Material online), declining toward the 3′-end of genes. LC samples showed the same overall methylation patterns (data not shown). This general pattern of DNA methylation appears consistent across social insect species (Glastad et al. 2017).

Individual Variability in DNA Methylation Profiles

We first investigated the overlap of mCGs across the individual LC samples. A total of 14,323 mCGs were found across the three samples after controlling for coverage differences (supplementary fig. S6*A*, Supplementary Material online). Of these, 6,312 mCGs (44.1%) were found in all samples. Conversely, 4,345 mCGs (30.3%) were found in one sample only (i.e., either in Sperm i3, i4, or i7), with the remaining 3,666 mCGs (25.6%) occurring in two of the three samples. In each individual LC semen sample, average methylation levels at each mCG site were close to 100% (average ∼96%; supplementary fig. S6*B* and *C*, Supplementary Material online), suggesting that the methylation status of all spermatozoa produced by an individual male is consistent (assuming no pseudoreplication at the read level).

HC samples also revealed individual variability in DNA methylation profiles. There were a total of 137,467 mCGs across all four samples after controlling for coverage differences (fig. 1*A*). Of these, 70,838 mCGs (51.5%) were found in all samples, 28,860 (21.0%) were found in one sample only (i.e., either in Sperm i47, i49, i51, or i53), with the remaining 37,769 mCGs (27.5%) occurring in two or three samples. Average levels of methylation were slightly lower in HC

Table 1
CG Methylation in the Four HC Semen Samples

Sample	Sperm i47	Sperm i49	Sperm i51	Sperm i53
No. sufficiently covered Cs (≥10 reads)	12,203,921	11,719,358	10,577,502	11,977,780
No. significantly methylated Cs (%)	126,490 (1.04%)	124,335 (1.06%)	122,423 (1.16%)	127,050 (1.06%)
No. shared significantly methylated Cs	70,838	70,838	70,838	70,838
Methylation level significantly methylated Cs (mean ± SE)	94.92 ± 0.03%	94.30 ± 0.03%	94.01 ± 0.03%	94.61 ± 0.03%
Coverage significantly methylated Cs (mean ± SE)	17.86 ± 0.04	17.39 ± 0.03	17.38 ± 0.04	17.74 ± 0.04

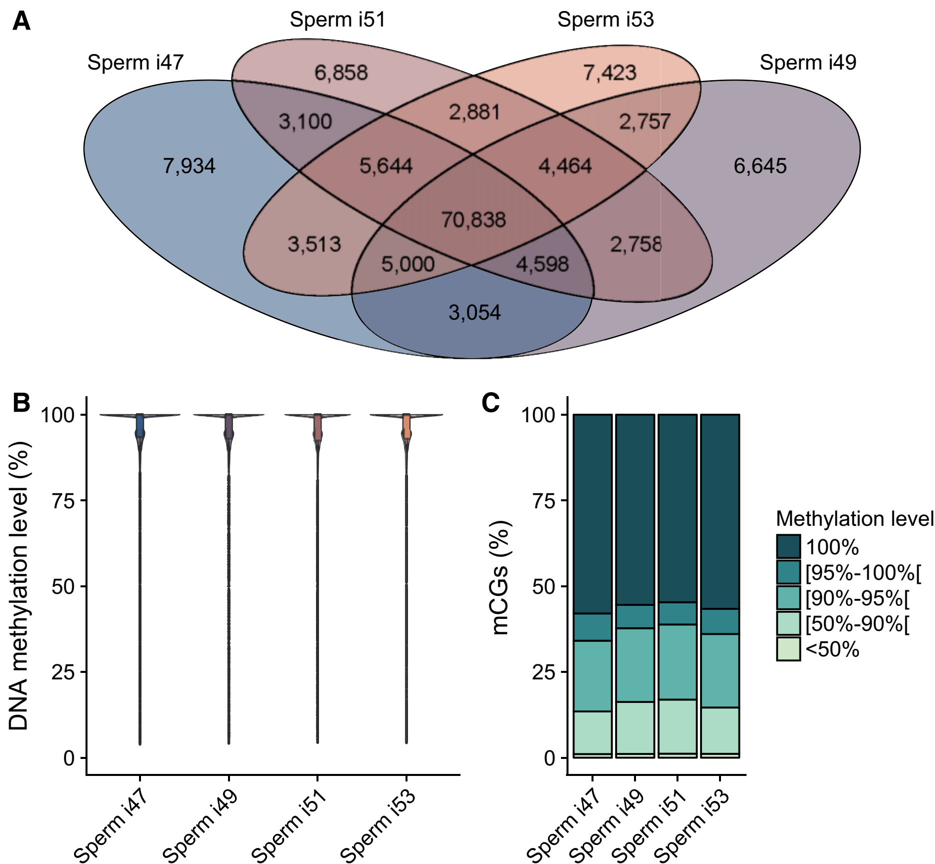


Fig. 1.—Individual variability and consistency in DNA methylation in honey bee semen. (*A*) Overlap of mCGs across the four individual HC semen samples. (*B*) Methylation level (%) of mCGs in each individual semen sample. (*C*) Proportion of mCGs showing methylation levels of respectively 100%, [95–100%], [90–95%], [50–90%], and <50% in each individual semen sample. Violin plots represent median, interquartile range, 95% confidence interval and kernel density plot.

samples compared with LC samples, suggesting a slight overestimation in LC samples. Yet, methylation levels of mCG sites were very high (~94%) and indicative of a largely consistent methylation status in all spermatozoa within each individual male (fig. 1*B* and *C*).

Differentially and Uniquely Methylated Genes

The number of methylated genes (defined as containing at least four mCGs and with a methylation level of at least 5%, supplementary fig. S1, Supplementary Material online) was 1,164–1,501 in LC samples, and 4,977–5,017 in HC samples. In HC samples, methylated genes represented ~38% of the genes covered from the honey bee genome, and accounted for ~83% of all mCGs. There was a core set of 4,693 methylated genes in all four HC samples, and 44–61 genes that were uniquely methylated in each individual semen sample (supplementary fig. S7, Supplementary Material online). Methylated genes in the individual semen samples had over 20 mCGs on average, spanning four exons and had genic

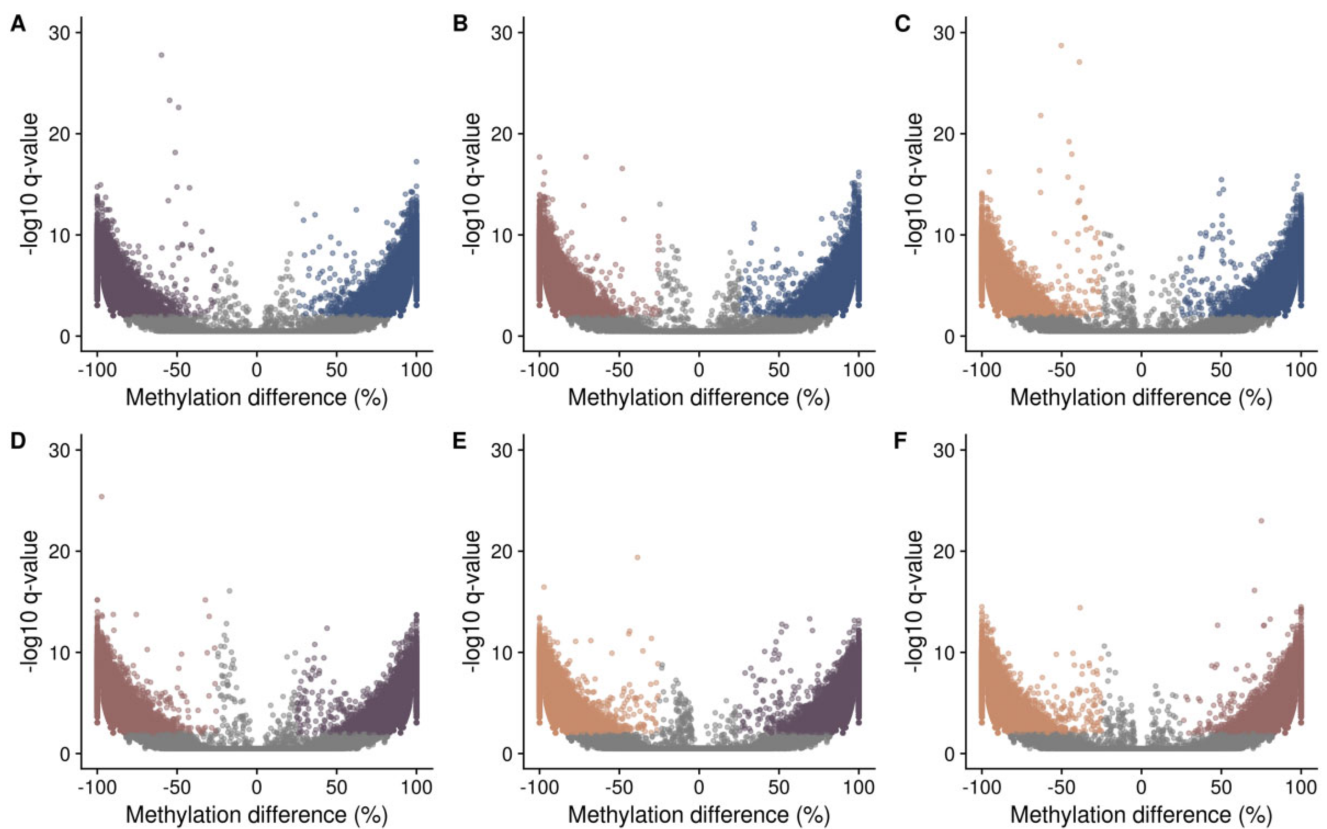


Fig. 2.—Differential methylation in male honey bees' semen. Volcano plots of significantly differentially methylated CGs in pairwise comparisons (Fisher's exact tests, *q*-value ≤0.01 and methylation difference >25%) between the four individual HC semen samples. (*A*) Sperm i47 (blue) versus Sperm i49 (purple). (*B*) Sperm i47 (blue) versus Sperm i51 (brown). (*C*) Sperm i47 (blue) versus Sperm i53 (orange). (*D*) Sperm i49 (purple) versus Sperm i51 (brown). (*E*) Sperm i49 (purple) versus Sperm i53 (orange). (*F*) Sperm i51 (brown) versus Sperm i53 (orange). Nonsignificant sites are represented in gray.

methylation levels of ~28%. Gene ontology analysis of the 4,693 CMGs in the four individual samples revealed an enrichment in constitutively expressed genes (supplementary table S8, Supplementary Material online), as is typically reported in other social insect species (Glastad et al. 2017).

To assess the between-sample variability in DNA methylation at the genic level, we performed pairwise comparisons to identify DMCs and infer DMGs (Remnant et al. 2016; Wang et al. 2016). There were between 267 and 405 DMCs across the three pairwise comparisons for LC samples (supplementary fig. S8, Supplementary Material online). Across pairwise comparisons the number of DMGs ranged from 194 to 270 (supplementary table S9, Supplementary Material online). The average number of DMCs per DMG was approximately 1.4, with ~25% of DMGs possessing more than one DMC (supplementary table S9, Supplementary Material online). There were between 37,086 and 38,729 DMCs across all six pairwise comparisons for HC samples (fig. 2). The number of DMGs was 4,394–4,545, each having on average 7.6 DMCs (table 2). Out of a total of 5,435 unique DMGs found when considering all pairwise comparisons together, 2,776 (51%) were consistently present in each pairwise comparison. Approximately 26% of DMGs (1,127–1,183 DMGs) contained at least ten DMCs and were retained for downstream analyses. Out of the 1,908 unique DMGs containing at least ten DMCs when considering all pairwise comparisons together, 492 (26%) were differentially methylated in each pairwise comparison.

Association between DNA Methylation and Sequence Polymorphism

To investigate the link between genetic variation and DNA methylation at the whole-genome level, we determined the presence of SNPs in all individual semen samples. Our SNP calling is likely underestimated because bisulfite treatment removes C–T polymorphisms, and because of the high rates of spontaneous deamination of methylated cytosines to thymines (Shen et al. 1994). We identified 44,569–49,671 SNPs in LC samples, and 134,222–142,001 SNPs in HC samples. As expected, SNPs were mostly located in intergenic regions (45–46%) and in introns (38–44%; supplementary tables S10 and

Table 2

Differentially Methylated Genes between the Four Individual HC Semen Samples

Pairwise Comparison	No. Hypermethylated DMGs	Total DMGs[a]	No. DMCs (mean ± SE)	No. DMGs with >10 DMCs
Sperm i47 vs. Sperm i49	3,571	4,520	7.69 ± 0.13	1,183 (26.17%)
Sperm i49 vs. Sperm i47	3,108			
Sperm i47 vs. Sperm i51	3,101	4,394	7.67 ± 0.13	1,145 (26.06%)
Sperm i51 vs. Sperm i47	3,389			
Sperm i47 vs. Sperm i53	3,267	4,507	7.62 ± 0.12	1,167 (25.89%)
Sperm i53 vs. Sperm i47	3,410			
Sperm i49 vs. Sperm i51	3,376	4,491	7.37 ± 0.12	1,127 (25.09%)
Sperm i51 vs. Sperm i49	3,274			
Sperm i49 vs. Sperm i53	3,502	4,545	7.65 ± 0.12	1,170 (25.74%)
Sperm i53 vs. Sperm i49	3,225			
Sperm i51 vs. Sperm i53	3,364	4,451	7.49 ± 0.12	1,168 (26.24%)
Sperm i53 vs. Sperm i51	3,223			

[a]Many genes contained hypermethylated sites in both samples.

S11, Supplementary Material online). To estimate the importance of sequence-specific methylation, we then focused on the SNPs located in *cis* (i.e., within 1 kb upstream or downstream) of an mCG. We identified between 728 and 979 *cis*-SNPs per LC sample (supplementary fig. S2 and table S12, Supplementary Material online), and between 4,170 and 4,515 *cis*-SNPs per HC sample (supplementary fig. S3; supplementary table S13, Supplementary Material online). In LC samples, *cis*-SNPs were found in 396–492 genes, and encompassed ~9% of all mCGs. In HC samples, *cis*-SNPs were found in 2,252–2,346 genes, and encompassed ~24% of all mCGs. Each mCG had on average 1.4 *cis*-SNPs within 1 kb across all semen samples. Conversely, each *cis*-SNP had on average ~3 mCGs within 1 kb in LC samples, and ~9 mCGs within 1 kb in HC samples. There was no obvious difference between B to A *cis*-SNPs (where B is C, G, or T) and H to G *cis*-SNPs (where H is A, C, or T) in their respective proportion and in the methylation level of their associated mCGs in the three LC semen samples (supplementary table S14, Supplementary Material online). In HC samples, B to A *cis*-SNPs were more frequent than H to G *cis*-SNPs, however the methylation level of their associated mCGs was identical (supplementary table S15, Supplementary Material online).

We first investigated the link between sequence polymorphism and differential methylation in LC samples. Despite the low number of *cis*-SNPs available in these samples, we found that, at the gene level, DMGs were more likely to contain *cis*-SNPs than genes that were consistently methylated (CMGs; χ^2 tests: all $P < 0.00001$; supplementary fig. S9*A–C*, Supplementary Material online). Further, DMGs had more *cis*-SNPs on average than CMGs (Wilcoxon rank sum tests: all $P < 0.0063$; supplementary fig. S9*D–F*, Supplementary Material online). This difference was not biased by gene length, which was not significantly different between DMGs and CMGs (all $P > 0.585$).

The association between sequence polymorphism and differential methylation was then considered in HC samples. In these four individual samples, DMGs also had a much higher propensity of containing *cis*-SNPs compared with CMGs (χ^2 tests: all $P < 0.00001$; fig. 3*A–D*). DMGs also had more *cis*-SNPs on average than CMGs (Wilcoxon rank sum tests: all $P < 0.00001$; fig. 3*E–H*). DMGs and CMGs also had similar length (all $P > 0.065$). We further found that methylated genes with *cis*-SNPs had more mCGs than methylated genes without *cis*-SNPs (Wilcoxon rank sum tests: all $P < 0.00001$; supplementary fig. S10*A–D*, Supplementary Material online). Methylated genes with *cis*-SNPs also had more DMCs (all $P < 0.00001$; supplementary fig. S10*E–H*, Supplementary Material online), and consequently lower methylation levels on average than methylated genes without *cis*-SNPs (Kruskal–Wallis rank sum tests: all $P < 0.0006$; supplementary fig. S10*I–L*, Supplementary Material online). Methylated genes with *cis*-SNPs were longer than methylated genes without *cis*-SNPs in Sperm i51 ($P = 0.032$), but not in the other three samples (all $P > 0.070$).

Direct Bisulfite PCR Sequencing of DMGs Confirms the Relationship between Sequence and Methylation

We performed direct bisulfite PCR sequencing of six DMGs in the semen of eight additional, unrelated males (fig. 4; supplementary fig. S11, Supplementary Material online). These genes were chosen based on their number of DMCs (supplementary table S16, Supplementary Material online), their *cis*-SNP density and the ability to design primers within regions containing mCGs and *cis*-SNPs (supplementary table S5, Supplementary Material online). To assign each individual semen sample to a particular allele, we also performed direct PCR sequencing of nonbisulfite converted DNA for all six genes. Each gene had between 2 and 6 alleles characterized by distinct sequence polymorphisms, including SNPs and small insertions and deletions (fig. 4; supplementary fig. S11; supplementary tables S17–S22, Supplementary Material online). *Oxysterol-binding protein-related protein 2* (GB52517) had

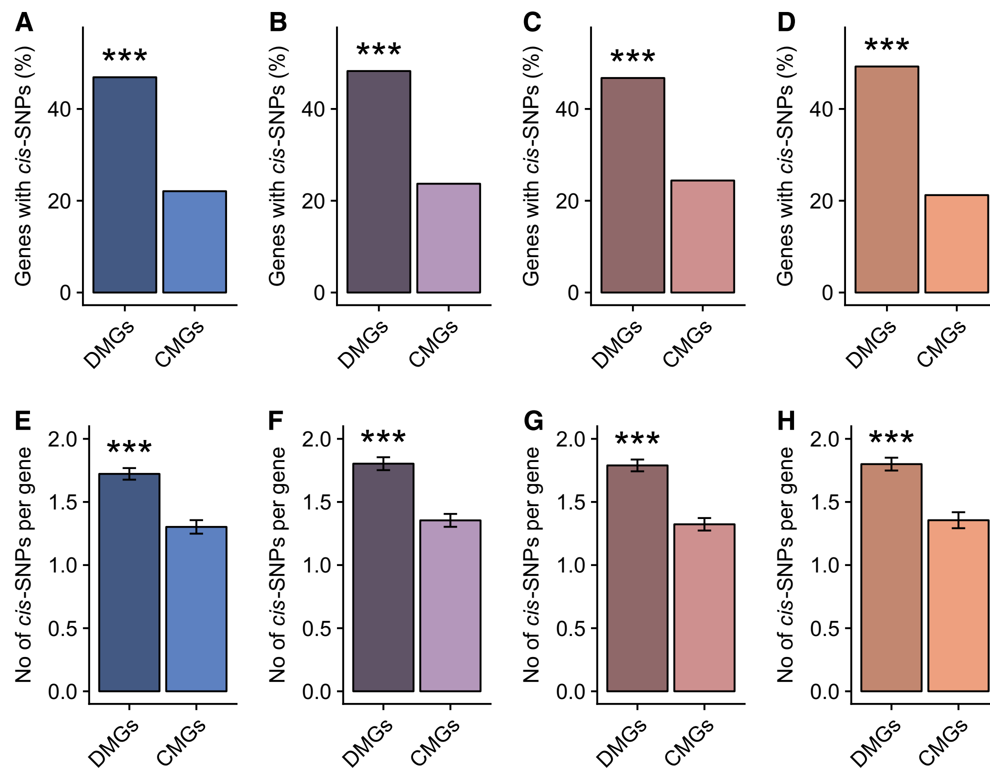


Fig. 3.—Differentially methylated genes (DMGs) have a higher tendency to exhibit sequence polymorphisms compared with consistently methylated genes (CMGs). (*A*, *E*) Sperm i47. (*B*, *F*) Sperm i49. (*C*, *G*) Sperm i51. (*D*, *H*) Sperm i53. Bar plots in (*E*–*H*) represent mean ± SE. (*A*–*D*): χ^2 tests: Sperm i47, $\chi^2 = 136.9$, df (degrees of freedom) = 1, $P < 0.00001$; Sperm i49, $\chi^2 = 135.6$, df = 1, $P < 0.00001$; Sperm i51, $\chi^2 = 111.0$, df = 1, $P < 0.00001$; Sperm i53, $\chi^2 = 172.4$, df = 1, $P < 0.00001$. (*E*–*H*): Wilcoxon rank sum tests: Sperm i47, $W = 81500$, $P < 0.00001$; Sperm i49, $W = 91156$, $P < 0.00001$; Sperm i51, $W = 90308$, $P < 0.00001$; Sperm i53, $W = 82917$, $P < 0.00001$. ***$P < 0.00001$.

two alleles (fig. 4*A*; supplementary table S17, Supplementary Material online), with lower average methylation density for allele 1 (59.0%) compared with allele 2 (75.4%). This was in part caused by SNPs removing two CG sites in allele 1, preventing methylation at these sites. Conversely, additional CG sites were generated by SNPs at three sites in allele 1, although not all individuals carrying allele 1 exhibited methylation at these sites (fig. 4*A*).

The three alleles identified for *protocadherin-like wing polarity protein stan* (GB51276) varied in average methylation density, with similar densities for allele 1 (52.4%) and allele 2 (51.6%), but lower in allele 3 (29.0%; fig. 4*B*; supplementary table S18, Supplementary Material online). Although some differences were due to loss or gain of CG sites by SNPs, not all methylation variation could be directly attributed to CG-related SNPs. This pattern was also observed for *atrial natriuretic peptide-converting enzyme* (GB54775), where methylation density was on average marginally higher in allele 1 (28.4%) than in allele 2 (22.6%), though overall methylation differences were not specifically due to SNPs causing loss or gain of CG sites (fig. 4*C*; supplementary table S19, Supplementary Material online). Methylation densities in the other three genes examined varied from 0% to 72% (supplementary fig. S11, Supplementary Material online). *Glutamate receptor ionotropic, NMDA 3A* (GB41839) had two alleles; allele 1 was completely unmethylated, whereas allele 2 had an average methylation density of 17.3% (supplementary fig. S11*A* and table S20, Supplementary Material online). *Heat shock protein Hsp70Ab-like* (GB50609) had six alleles with methylation densities ranging from 0% to 57.7% (supplementary fig. S11*B* and table S21, Supplementary Material online). *Parafibromin* (GB43888) had four alleles with methylation densities ranging from 46.0% to 72.0% (supplementary fig. S11*C* and table S22, Supplementary Material online).

Sequence-specific methylation was evident for some CG sites, because of loss or gain of CG sites, and because of non-CG-related SNPs in *cis*. Further, there was a substantial within-allele variation in methylation patterns. Agglomerative hierarchical clustering of methylation patterns in all eight individuals, focusing on the four genes that had two or three allelic variants, showed that individuals having the same allele tended to cluster with respect to their methylation patterns (supplementary fig. S12, Supplementary Material online).

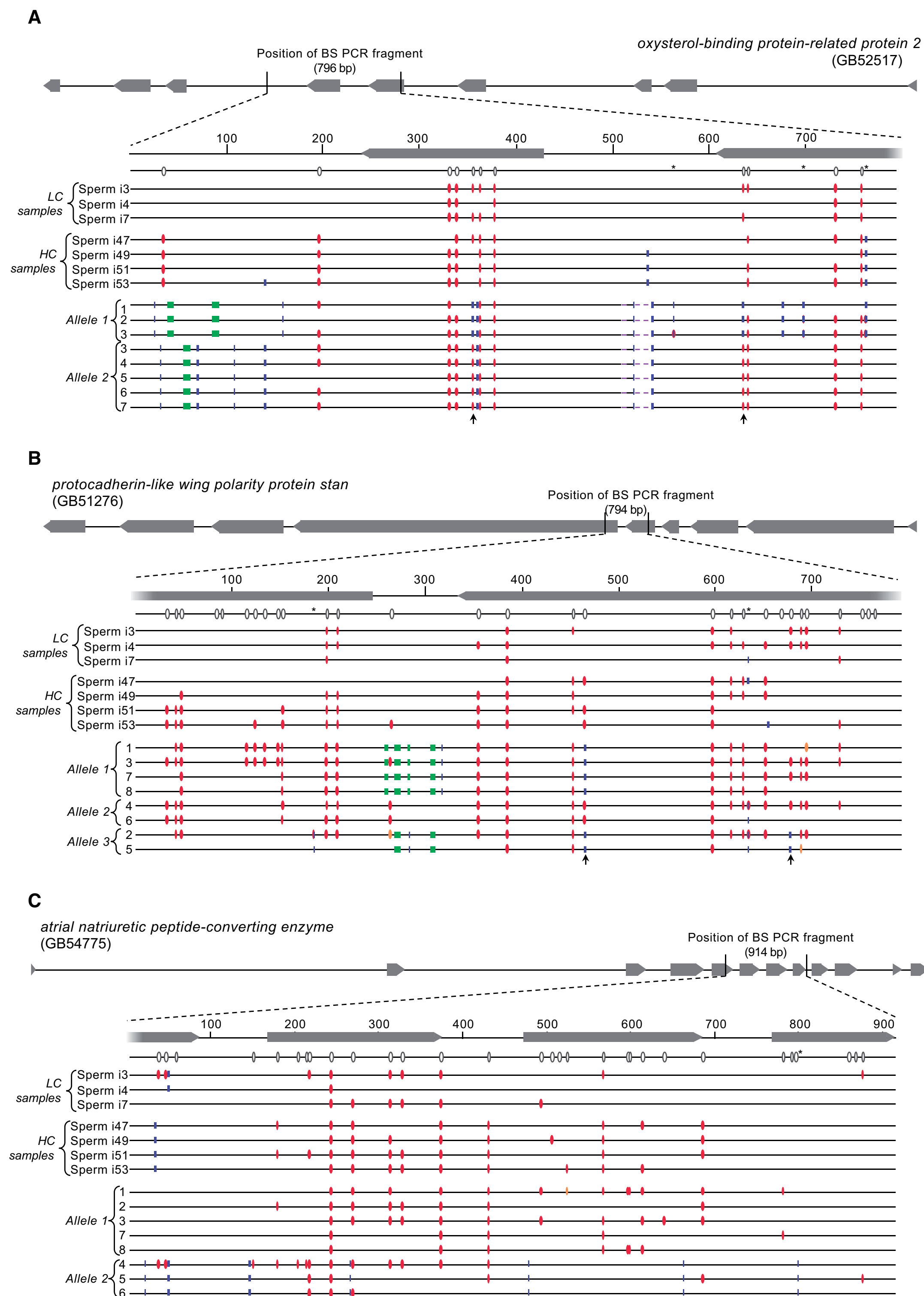


Fig. 4.—Direct bisulfite PCR sequencing of (*A*) *oxysterol-binding protein-related protein 2* (GB52517), (*B*) *protocadherin-like wing polarity protein stan* (GB51276), and (*C*) *atrial natriuretic peptide-converting enzyme* (GB54775). Exons are represented in gray. CG sites of the reference genome are represented by open ovals. WGBS LC and HC individual samples are represented. Lines labeled 1–8 are direct bisulfite PCR sequencing individual samples. mCGs are represented as red ovals. Orange ovals indicate where C/T peaks were present indicating sites with intermediate (<100%) methylation. SNPs are represented as blue rectangles. Insertions are represented as green polygons. Deletions are represented as purple dashed lines. Stars indicate SNPs adding CG sites absent from the reference genome. Arrows indicate SNPs removing CG sites present in the reference genome.

Discussion

We show that honey bee males harbor individual-specific DNA methylation patterns in their semen and that these patterns are often associated with genotypic variation. Each semen sample possessed a unique methylation profile that was associated with sequence polymorphisms in *cis*. Genes that are variable at the epigenetic level are more likely to be variable at the genetic level. We propose that sequence polymorphism could be an important determinant of DNA methylation state at many loci in honey bees, contributing both to the individual-specificity of epigenetic marks, and to their retention across generations.

In this study, we show that genetically different individuals harbor different methylomes, however it remains to be seen whether genetically identical individuals harbor identical methylomes. Whether genetic variation causally determines DNA methylation states, or alternatively whether genetic and epigenetic variations covary due to *trans*-acting factors (Tycko 2010) remains an open question. In honey bees, because constitutively expressed genes tend to be both more methylated and evolutionarily conserved (Hunt et al. 2013), this could also help explain the association between genetic and epigenetic variation. Although the relationship between DNA methylation and gene expression may not be as straightforward as originally thought (e.g., Bewick et al. 2019), investigating gene expression together with DNA methylation across generations would provide valuable insights as to the prevalence of this nonmutually exclusive mechanism.

The role of *cis*-mediated allele-specific methylation in epigenetic inheritance is emerging as a key factor in understanding variability in human polymorphisms (Kerkel et al. 2008; Tycko 2010; Heard and Martienssen 2014; Mendizabal et al. 2014). *Cis*-mediated allele-specific methylation has also been observed in a solitary insect (Wang et al. 2016). Further, there is clear evidence of an effective transfer of DNA methylation marks from fathers to offspring in several species (Rando 2016; Tillo et al. 2016; Skvortsova et al. 2019). In mammals, genome-wide DNA methylation reprograming is a feature of gametogenesis (Wang et al. 2014), however, fathers that are exposed to environmental stressors acquire altered methylation profiles in their sperm, which can be transmitted to subsequent generations (Anway et al. 2005; Radford et al. 2014). Inheritance of sperm-specific methylation appears to be mediated by genetic variation between individuals (Heard and Martienssen 2014; Shea et al. 2015). In zebrafish, the sperm methylome is stably transmitted to the embryo, whereas the oocyte methylome undergoes extensive reprograming during embryogenesis (Jiang et al. 2013; Skvortsova et al. 2019). In *Nasonia* wasps there is stable inheritance of methylation states between generations, likely under *cis*-regulation (Wang et al. 2016). Whether stable germline transmission of DNA methylation occurs in honey bees remains an open question. DNA methylation patterns are thought to change during gametogenesis and development, though current evidence suggests a lack of extensive reprograming (Drewell et al. 2014; Maleszka 2016).

Overall, it seems possible that honey bee males transmit DNA methylation patterns through their sperm to their worker daughters, particularly if, as we show, this methylation state is associated with DNA sequence variation. We find that the methylation level of each mCG is largely consistent, indicating that spermatozoa within an individual's ejaculate have the same methylation status. This intraindividual homogeneity and interindividual heterogeneity is consistent with what has been reported in mammals (Jenkins et al. 2015; Dere et al. 2016). In the honey bee, if these epigenetic marks are transmitted across generations, this would create specific DNA methylation profiles in a particular male's worker daughters that would differ from those of another male's daughters because all the workers sired by a particular male inherit a similar methylome. Worker phenotypes differ in a patriline-dependent manner in honey bees (reviewed by Oldroyd and Fewell 2007) and other social insects (e.g., Julian and Fewell 2004). Generally these differences are held to have a genetic origin (Smith et al. 2008). However, heritable epigenetic influences on patriline-specific worker traits could also be important, and probably complementary to genetic differences.

The possibility that honey bee males epigenetically influence their worker progeny through DNA methylation could be relevant in the context of genomic imprinting (Oldroyd et al. 2014; Galbraith et al. 2016). Recent findings suggest that epigenetic mechanisms underlie parent-of-origin effects relating to worker reproduction. First, queens and males produce gametes with distinct methylomes (Drewell et al. 2014). Second, methylation profiles of sexually-produced embryos (inheriting a maternal and paternal genome) have increased methylation relative to parthenogenetically-produced embryos (inheriting only a maternal genome) (Remnant et al. 2016). Together, these observations suggest that male and female gametes each contribute a distinct methylation profile to offspring. Third, we have shown here that individual males contribute unique methylation profiles to offspring, as they produce semen with specific DNA methylation profiles that are linked to genetic variation. Many genes were consistently methylated across all semen samples. Among these there could be loci that are methylated with male-specific patterns during spermatogenesis, and that are transmitted to female offspring. Testing this hypothesis would require comparing the methylation profile of these genes in female oocytes and their expression levels in workers.

We note that other epigenetic mechanisms, such as small RNAs (in particular Piwi-interacting RNAs) and histone post-translational modifications, also play important roles in regulating gene function (Suganuma and Workman 2011; Erwin et al. 2015). Both mechanisms are associated with DNA methylation (Aravin et al. 2008; Cedar and Bergman 2009) and are

demonstrated phenomena in honey bees (Ashby et al. 2016; Wojciechowski et al. 2018).

Sequence-specific methylation at a genome-wide level raises important methodological questions about quantifying methylation in DNA extracted from pools of individuals. Unless the sources of DNA are individual haploids, a degree of genetic heterogeneity will always be encountered within samples. Small sample sizes and/or replicate numbers then risk unequal representation of alleles within pools. When this happens one might erroneously conclude that two pools are differentially methylated as a result of experimental manipulation, whereas in reality the impression of differential methylation is a consequence of genomic variation in samples (Libbrecht et al. 2016), which we show here can lead to sequence-specific methylation. This risk is particularly acute in polyandrous species like honey bees. If worker samples originate from different patrilines, then they will likely have different methylation patterns. If the DNA pools being compared comprise unequal patriline proportions, the potential for error is clear. A robust experimental design is therefore necessary to disentangle the influence of epigenetic modifications from underlying genetic variation.

Although high coverage and sequencing quality are necessary to fully grasp the extent of sequence-specific methylation and to reduce the prevalence of false positive findings, our LC and HC samples provided identical conclusions. This result indicates that low coverage is sufficient to generate robust results using genetically homogeneous samples. However, in genetically heterogeneous samples, biases such as allelic skew may be introduced by low coverage if multiple alleles are present and not all are equally represented.

We conclude that the interplay of genetic and epigenetic factors must be considered when interpreting methylation studies. For some methylated sites there may be a direct causative link between genotype and epigenetic mark. In others there may be complete independence of epigenetic and genetic variation (Miska and Ferguson-Smith 2016). In between, many sites are likely to show a probabilistic relationship between genotype and epigenotype, that is, the presence of epigenetic marks is influenced by genotype, but also affected by other environmental or context-dependent factors (Richards 2006). Whether deterministically or probabilistically, if genotype and methylation state are linked, then epigenetic patterns are subject to selection and could thus be adaptive.

Supplementary Material

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

Acknowledgments

We thank Stephen Rose for his help with bioinformatics analyses. This work was supported by the Australian Research Council [DP150100151 to B.P.O. and Alyson Ashe]; the School of Life and Environmental Sciences at the University of Sydney, the Fyssen Foundation [to B.Y.]; and the Hermon Slade Foundation [HSF1801 to B.P.O, B.Y., and E.J.R.]. Library preparation for WGBS was performed at the Westmead Scientific Platforms, which are supported by the Westmead Research Hub, the Westmead Institute for Medical Research, the Cancer Institute New South Wales, the National Health and Medical Research Council, and the Ian Potter Foundation.

Literature Cited

Akalin A, et al. 2012. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biol. 13(10):R87.

Anway MD, Cupp AS, Uzumcu M, Skinner MK. 2005. Epigenetic transgenerational actions of endocrine disruptors and male fertility. Science 308(5727):1466–1469.

Aravin AA, et al. 2008. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Mol Cell 31(6):785–799.

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.

Ashby R, Foret S, Searle I, Maleszka R. 2016. MicroRNAs in honey bee caste determination. Sci Rep. 6:18794.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate—a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol. 57:289–300.

Berger SL, Kouzarides T, Shiekhattar R, Shilatifard A. 2009. An operational definition of epigenetics. Genes Dev. 23(7):781–783.

Bewick AJ, Zhang Y, Wendte JM, Zhang X, Schmitz RJ. 2019. Evolutionary and experimental loss of gene body methylation and its consequence to gene expression. G3 (Bethesda). 9(8):2441–2445.

Bonasio R. 2014. The role of chromatin and epigenetics in the polyphenisms of ant castes. Brief Funct Genomics 13(3):235–245.

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.

Bonasio R, Tu S, Reinberg D. 2010. Molecular signals of epigenetic states. Science 330(6004):612–616.

Cedar H, Bergman Y. 2009. Linking DNA methylation and histone modification: patterns and paradigms. Nat Rev Genet. 10(5):295–304.

Cridge A, Harrop T, Lovegrove M, Remnant EJ, Dearden P. 2017. Nutrition and epigenetic change in insects: evidence and implications. Chap 2. In: Verlinden H, editor. Advances in insect physiology. Amsterdam: Elsevier, Vol. 53. p. 31–54.

Danecek P, et al. 2011. The variant call format and VCFtools. Bioinformatics 27(15):2156–2158.

Dere E, Huse S, Hwang K, Sigman M, Boekelheide K. 2016. Intra- and inter-individual differences in human sperm DNA methylation. Andrology 4(5):832–842.

Do C, et al. 2017. Genetic-epigenetic interactions in *cis*: a major focus in the post-GWAS era. Genome Biol. 18(1):120.

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.

Drewell RA, Lo N, Oxley PR, Oldroyd BP. 2012. Kin conflict in insect societies: a new epigenetic perspective. Trends Ecol Evol. 27(7):367–373.

Elsik CG, et al. 2014. Finding the missing honey bee genes: lessons learned from a genome upgrade. BMC Genomics 15:86.

Erwin AA, et al. 2015. piRNAs are associated with diverse transgenerational effects on gene and transposon expression in a hybrid dysgenic syndrome of *D. virilis*. PLoS Genet. 11(8):e1005332.

Fang F, et al. 2012. Genomic landscape of human allele-specific DNA methylation. Proc Natl Acad Sci U S A. 109(19):7332–7337.

Galbraith DA, 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.

Glastad KM, et al. 2017. Variation in DNA methylation is not consistently reflected by sociality in Hymenoptera. Genome Biol Evol. 9(6):1687–1698.

Glastad KM, Chau LM, Goodisman MAD. 2015. Epigenetics in social insects. In: Zayed A, Kent CF, editors. Advance in insect physiology. Amsterdam: Elsevier, Vol. 48. p. 227–269.

Glastad KM, Hunt BG, Goodisman MAD. 2014. Evolutionary insights into DNA methylation in insects. Curr Opin Insect Sci. 1:25–30.

Glastad KM, Hunt BG, Goodisman MAD. 2019. Epigenetics in insects: genome regulation and the generation of phenotypic diversity. Annu Rev Entomol. 64:185–203.

Glastad KM, Hunt BG, Yi SV, Goodisman MA. 2011. DNA methylation in insects: on the brink of the epigenomic era. Insect Mol Biol. 20(5):553–565.

Goldberg AD, Allis CD, Bernstein E. 2007. Epigenetics: a landscape takes shape. Cell 128(4):635–638.

Guzman-Novoa E, et al. 2005. Paternal effects on the defensive behavior of honeybees. J Hered. 96(4):376–380.

Haig D. 2000. The kinship theory of genomic imprinting. Annu Rev Ecol Syst. 31(1):9–32.

Heard E, Martienssen RA. 2014. Transgenerational epigenetic inheritance: myths and mechanisms. Cell 157(1):95–109.

Heinz S, et al. 2010. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. Mol Cell 38(4):576–589.

Herb BR, et al. 2012. Reversible switching between epigenetic states in honeybee behavioral subcastes. Nat Neurosci. 15(10):1371–1373.

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.

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.

Jeltsch A, Jurkowska RZ. 2014. New concepts in DNA methylation. Trends Biochem Sci. 39(7):310–318.

Jenkins TG, et al. 2015. Intra-sample heterogeneity of sperm DNA methylation. Mol Hum Reprod. 21(4):313–319.

Jiang L, et al. 2013. Sperm, but not oocyte, DNA methylome is inherited by zebrafish early embryos. Cell 153(4):773–784.

Julian GE, Fewell JH. 2004. Genetic variation and task specialization in the desert leaf-cutter ant, *Acromyrmex versicolor*. Anim Behav. 68(1):1–8.

Kearse M, et al. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12):1647–1649.

Kerkel K, et al. 2008. Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. Nat Genet. 40(7):904–908.

Klose RJ, Bird AP. 2006. Genomic DNA methylation: the mark and its mediators. Trends Biochem Sci. 31(2):89–97.

Kocher SD, et al. 2015. A search for parent-of-origin effects on honey bee gene expression. G3 (Bethesda) 5(8):1657–1662.

Koeniger G, Koeniger N, Ellis J, Connor L. 2014. Mating biology of honey bees (*Apis mellifera*). Kalamazoo (MI): Wicwas Press.

Kohli RM, Zhang Y. 2013. TET enzymes, TDG and the dynamics of DNA demethylation. Nature 502(7472):472–479.

Krueger F, Andrews SR. 2011. Bismark: a flexible aligner and methylation caller for bisulfite-seq applications. Bioinformatics 27(11):1571–1572.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9(4):357–359.

Li-Byarlay H. 2016. The function of DNA methylation marks in social insects. Front Ecol Evol. 4:57.

Li-Byarlay H, et al. 2013. RNA interference knockdown of *DNA methyltransferase 3* affects gene alternative splicing in the honey bee. Proc Natl Acad Sci U S A. 110(31):12750–12755.

Libbrecht R, Oxley PR, Keller L, Kronauer DJ. 2016. Robust DNA methylation in the clonal raider ant brain. Curr Biol. 26(3):391–395.

Lienert F, et al. 2011. Identification of genetic elements that autonomously determine DNA methylation states. Nat Genet. 43(11):1091–1097.

Lyko F, et al. 2010. The honey bee epigenomes: differential methylation of brain DNA in queens and workers. PLoS Biol. 8(11):e1000506.

Maleszka R. 2016. Epigenetic code and insect behavioural plasticity. Curr Opin Insect Sci. 15:45–52.

Meaburn EL, Schalkwyk LC, Mill J. 2010. Allele-specific methylation in the human genome: implications for genetic studies of complex disease. Epigenetics 5(7):578–582.

Mendizabal I, Keller TE, Zeng J, Yi SV. 2014. Epigenetics and evolution. Integr Comp Biol. 54(1):31–42.

Miska EA, Ferguson-Smith AC. 2016. Transgenerational inheritance: models and mechanisms of non-DNA sequence-based inheritance. Science 354(6308):59–63.

Oldroyd BP, et al. 2014. A parent-of-origin effect on honeybee worker ovary size. Proc Biol Sci. 281(1775):20132388.

Oldroyd BP, Fewell JH. 2007. Genetic diversity promotes homeostasis in insect colonies. Trends Ecol Evol. 22(8):408–413.

Oster GF, Wilson EO. 1978. Caste and ecology in the social insects. Princeton (NJ): Princeton University Press.

Palmer K, Oldroyd BP. 2000. Evolution of multiple mating in the genus *Apis*. Apidologie 31(2):235–248.

Queller DC. 2003. Theory of genomic imprinting conflict in social insects. BMC Evol Biol. 3:15.

R Core Team. 2017. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Radford EJ, et al. 2014. In utero undernourishment perturbs the adult sperm methylome and intergenerational metabolism. Science 345(6198):1255903.

Rando OJ. 2016. Intergenerational transfer of epigenetic information in sperm. Cold Spring Harb Perspect Med. 6(5):a022988.

Ratel D, Ravanat JL, Berger F, Wion D. 2006. N6-methyladenine: the other methylated base of DNA. Bioessays 28(3):309–315.

Reid RJ, Remnant EJ, Allsopp MH, Beekman M, Oldroyd BP. 2017. Paternal effects on *Apis mellifera capensis* worker ovary size. Apidologie 48(5):660–665.

Reik W, Walter J. 2001. Genomic imprinting: parental influence on the genome. Nat Rev Genet. 2(1):21–32.

Remnant EJ, et al. 2016. Parent-of-origin effects on genome-wide DNA methylation in the Cape honey bee (*Apis mellifera capensis*) may be confounded by allele-specific methylation. BMC Genomics 17:226.

Richards EJ. 2006. Inherited epigenetic variation—revisiting soft inheritance. Nat Rev Genet. 7(5):395–401.

Sambrook J, Fritsch EF, Maniatis T. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

Schilling E, El Chartouni C, Rehli M. 2009. Allele-specific DNA methylation in mouse strains is mainly determined by *cis*-acting sequences. Genome Res. 19(11):2028–2035.

Schwander T, Lo N, Beekman M, Oldroyd BP, Keller L. 2010. Nature versus nurture in social insect caste differentiation. Trends Ecol Evol. 25(5):275–282.

Shea JM, et al. 2015. Genetic and epigenetic variation, but not diet, shape the sperm methylome. Dev Cell 35(6):750–758.

Shen JC, Rideout IIW, Jones PA. 1994. The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. Nucleic Acids Res. 22(6):972–976.

Shoemaker R, Deng J, Wang W, Zhang K. 2010. Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. Genome Res. 20(7):883–889.

Simpson SJ, Sword GA, Lo N. 2011. Polyphenism in insects. Curr Biol. 21(18):R738–R749.

Skvortsova K, et al. 2019. Retention of paternal DNA methylome in the developing zebrafish germline. Nat Commun. 10(1):3054.

Smith CR, et al. 2012. Patterns of DNA methylation in development, division of labor and hybridization in an ant with genetic caste determination. PLoS One 7(8):e42433.

Smith CR, Toth AL, Suarez AV, Robinson GE. 2008. Genetic and genomic analyses of the division of labour in insect societies. Nat Rev Genet. 9(10):735–748.

Standage DS, et al. 2016. Genome, transcriptome and methylome sequencing of a primitively eusocial wasp reveal a greatly reduced DNA methylation system in a social insect. Mol Ecol. 25(8):1769–1784.

Suganuma T, Workman JL. 2011. Signals and combinatorial functions of histone modifications. Annu Rev Biochem. 80:473–499.

Suzuki MM, Bird A. 2008. DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet. 9(6):465–476.

Suzuki R, Shimodaira H. 2006. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics 22(12):1540–1542.

Terrapon N, et al. 2014. Molecular traces of alternative social organization in a termite genome. Nat Commun. 5(1):3636.

The Honeybee Genome Sequencing Consortium. 2006. Insights into social insects from the genome of the honeybee *Apis mellifera*. Nature 443:931–949.

Tillo D, Mukherjee S, Vinson C. 2016. Inheritance of cytosine methylation. J Cell Physiol. 231(11):2346–2352.

Tycko B. 2010. Allele-specific DNA methylation: beyond imprinting. Hum Mol Genet. 19(R2):R210–R220.

Wang H-Q, Tuominen LK, Tsai C-J. 2011. SLIM: a sliding linear model for estimating the proportion of true null hypotheses in datasets with dependence structures. Bioinformatics 27(2):225–231.

Wang L, et al. 2014. Programming and inheritance of parental DNA methylomes in mammals. Cell 157(4):979–991.

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.

Wedd L, Kucharski R, Maleszka R. 2016. Differentially methylated obligatory epialleles modulate context-dependent *LAM* gene expression in the honeybee *Apis mellifera*. Epigenetics 11(1):1–10.

Weiner SA, Toth AL. 2012. Epigenetics in social insects: a new direction for understanding the evolution of castes. Genet Res Int. 2012:609810.

Welch M, Lister R. 2014. Epigenomics and the control of fate, form and function in social insects. Curr Opin Insect Sci. 1:31–38.

West-Eberhard MJ. 2003. Developmental plasticity and evolution. New York: Oxford University Press.

Williams JL. 1987. Wind-directed pheromone trap for drone honey bees (Hymenoptera: Apidae). J Econ Entomol. 80(2):532–536.

Wojciechowski M, et al. 2018. Phenotypically distinct female castes in honey bees are defined by alternative chromatin states during larval development. Genome Res. 28(10):1532–1542.

Yagound B, Duncan M, Chapman NC, Oldroyd BP. 2017. Subfamily-dependent alternative reproductive strategies in worker honeybees. Mol Ecol. 26(24):6938–6947.

Yan H, et al. 2015. DNA methylation in social insects: how epigenetics can control behavior and longevity. Annu Rev Entomol. 60:435–452.

Zemach A, McDaniel IE, Silva P, Zilberman D. 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science 328(5980):916–919.

Associate editor: Maria Costantini