

Variation in DNA Methylation Is Not Consistently Reflected by Sociality in Hymenoptera

Karl M. Glastad¹, Samuel V. Arsenault², Kim L. Vertacnik³, Scott M. Geib⁴, Sasha Kay², Bryan N. Danforth⁵, Sandra M. Rehan⁶, Catherine R. Linnen³, Sarah D. Kocher^{7,*}, and Brendan G. Hunt^{2,*}

¹Department of Cell and Developmental Biology, University of Pennsylvania

²Department of Entomology, University of Georgia

³Department of Biology, University of Kentucky

⁴U.S. Department of Agriculture-Agricultural Research Service (USDA-ARS), Daniel K. Inouye U.S. Pacific Basin Agricultural Research Center, Hilo, Hawaii

⁵Department of Entomology, Cornell University

⁶Department of Biological Sciences, University of New Hampshire, Durham, New Hampshire

⁷Lewis-Sigler Institute for Integrative Genomics, Princeton University

*Corresponding authors: E-mails: huntbg@uga.edu; skocher@gmail.com.

Accepted: July 9, 2017

Data deposition: This project has been deposited at Gene Expression Omnibus under the accession GSE93893.

Abstract

Changes in gene regulation that underlie phenotypic evolution can be encoded directly in the DNA sequence or mediated by chromatin modifications such as DNA methylation. It has been hypothesized that the evolution of eusocial division of labor is associated with enhanced gene regulatory potential, which may include expansions in DNA methylation in the genomes of Hymenoptera (bees, ants, wasps, and sawflies). Recently, this hypothesis garnered support from analyses of a commonly used metric to estimate DNA methylation in silico, CpG content. Here, we test this hypothesis using direct, nucleotide-level measures of DNA methylation across nine species of Hymenoptera. In doing so, we generated new DNA methylomes for three species of interest, including one solitary and one facultatively eusocial halictid bee and a sawfly. We demonstrate that the strength of correlation between CpG content and DNA methylation varies widely among hymenopteran taxa, highlighting shortcomings in the utility of CpG content as a proxy for DNA methylation in comparative studies of taxa with sparse DNA methylomes. We observed strikingly high levels of DNA methylation in the sawfly relative to other investigated hymenopterans. Analyses of molecular evolution suggest the relatively distinct sawfly DNA methylome may be associated with positive selection on functional DNMT3 domains. Sawflies are an outgroup to all ants, bees, and wasps, and no sawfly species are eusocial. We find no evidence that either global expansions or variation within individual ortholog groups in DNA methylation are consistently associated with the evolution of social behavior.

Key words: CpG o/e, epigenetics, eusociality, *Neodiprion lecontei*, social insects, whole genome bisulfite sequencing.

Introduction

Epigenetic information influences phenotypes by stably altering chromosome structure (Berger et al. 2009). One form of epigenetic information is the methylation of DNA, which occurs primarily at cytosines in CpG dinucleotides in animal genomes (Goll and Bestor 2005). Although DNA methylation occurs globally in the genomes of vertebrates, it is primarily restricted to a subset of transcribed genes in the genomes of

investigated insects with functional DNA methylation systems (Suzuki and Bird 2008; Zemach et al. 2010; Glastad et al. 2011). DNA methylation has been shown to influence gene regulation by altering transcription factor binding (Wang et al. 2012), alternative splicing (Shukla et al. 2011; Li-Byarlay et al. 2013), and transcriptional elongation (Zilberman et al. 2007). DNA methylation can be altered during the course of development (Jaenisch and Bird 2003), and several studies of

eusocial insects suggest that DNA methylation may be capable of influencing developmental plasticity (Kucharski et al. 2008; Herb et al. 2012; Li-Byarlay et al. 2013; Alvarado et al. 2015; Glastad, Gokhale, et al. 2016; cf. Libbrecht et al. 2016).

Normalized CpG content (CpG o/e) can serve as a proxy measure for DNA methylation because methylated cytosines undergo deamination to thymine with high frequency (Shen et al. 1994; Elango et al. 2009). CpG o/e has been widely used in studies of diverse eukaryotes to gain insight into the genomic targets of DNA methylation (Yi and Goodisman 2009). More recently, CpG o/e was used to assess variation in DNA methylation among ten bee genomes (Kapheim et al. 2015). This analysis provided support for the hypothesis that evolutionary variation in the pervasiveness of DNA methylation is associated with taxonomic variation in social complexity among hymenopteran insects (but see Patalano et al. 2015; Standage et al. 2016; Bewick et al. 2017).

We analyzed whole genome bisulfite sequencing (WGBS) data from whole bodies of females from nine species in the insect order Hymenoptera to generate genome-wide, nucleotide-level measures of DNA methylation (DNA methylomes). This order includes sawflies, wasps, bees, and ants. To complement published data sets, we generated the first WGBS data for the sawfly *Neodiprion lecontei* and the bees *Lasioglossum albipes*, *Dufuorea novaeangliae*, and *Ceratina calcarata* (*C. calcarata* data previously reported in Rehan et al. 2016). The species we compare represent major groups within Hymenoptera and encompass two origins of intermediate social behavior (in halictid bees [facultatively eusocial] and xylocopine bees [subsocial with facultative alloparental care]) and two origins of advanced eusocial behavior (ants and honey bees; social terminology *sensu* Wilson 1971; Kocher and Paxton 2014). The primary goal of our study is to investigate the potential for convergence in patterns of DNA methylation with respect to social complexity in Hymenoptera, both globally and, for the first time, on a gene-by-gene basis. In doing so, we provide an assessment of the utility of CpG o/e in comparative analyses of taxa with sparse DNA methylomes, compare genomic levels and patterns of DNA methylation among taxa, and provide insight into the correlates of evolutionary variation in DNA methylation within ortholog groups. We also provide an examination of the molecular evolution of three enzymes that mediate DNA methylation in Hymenoptera, with the goal of identifying candidate mechanisms for evolutionary changes to DNA methylation patterns.

Several recent studies of DNA methylation in insect taxa have cast doubt on the proposed association between DNA methylation and reproductive division of labor (Patalano et al. 2015; Libbrecht et al. 2016; Standage et al. 2016; Bewick et al. 2017). However, our experimental design builds upon these findings in several important ways. First, we generated new DNA methylomes for several taxa of interest to the relationship between DNA methylation and social complexity.

Our study includes nine species from separate genera with comparable, complete sets of DNA methylation enzymes (DNMT1 and DNMT3). In contrast, five species of Hymenoptera with confirmed presence of both DNMT1 and DNMT3 and DNA methylomes are represented in Bewick et al. (2017), among which only one is not eusocial. Two other recent studies observed sparse DNA methylation in eusocial species, but these studies either did not incorporate analyses of noneusocial taxa (Patalano et al. 2015) or did not directly compare fractional levels of DNA methylation across taxa (Standage et al. 2016). Importantly, our sequencing depth facilitates analyses of variation in DNA methylation among members of individual ortholog groups, which enables a gene-specific test of convergent associations between DNA methylation and social complexity.

Results

Comparison of CpG Depletion and WGBS Data

We analyzed DNA methylation at CpG dinucleotides covered by ten or more reads in the reference genomes of nine hymenopteran species (56–98% of CpGs in each genome; supplementary tables S1 and S2, Supplementary Material online). We chose not to include recent DNA methylomes from *Polistes* aculeate wasps (Patalano et al. 2015; Standage et al. 2016) because *Polistes* are absent DNMT3, exhibit very little DNA methylation, and these data were not generated from whole bodies. We first compared our results to previously published work relying on CpG depletion to detect evolutionary variation in DNA methylation. DNA methylation has been observed primarily in exons in hymenopteran genomes (Wang et al. 2013; Bonasio et al. 2012; Hunt et al. 2013; Zemach et al. 2010; Rehan et al. 2016), so we compared the methylation status of coding sequences (according to WGBS data) to normalized CpG content of coding sequences (CpG o/e). DNA methylation must uniformly influence CpG substitution in distinct taxa for CpG o/e to provide a robust metric of evolutionary variation in DNA methylation. In the ideal case, we would expect similar bimodal distributions of CpG o/e in each taxon, with low CpG o/e values for methylated genes and high CpG o/e values for unmethylated genes (Elango et al. 2009).

We did observe that CpG o/e is clearly and significantly influenced by DNA methylation in all taxa in our study (supplementary figs. S1–S3, Supplementary Material online), but there was also striking variation among taxa in the extent to which CpG o/e distributions reflected DNA methylation status (fig. 1 and supplementary figs. S2–S4 and table S3, Supplementary Material online; Gadau et al. 2012; Glastad et al. 2011). For example, the vast majority of coding sequences targeted by DNA methylation in the bees *L. albipes* and *C. calcarata* were not discernable from unmethylated genes in terms of CpG content (fig. 1). Spearman's rank correlations

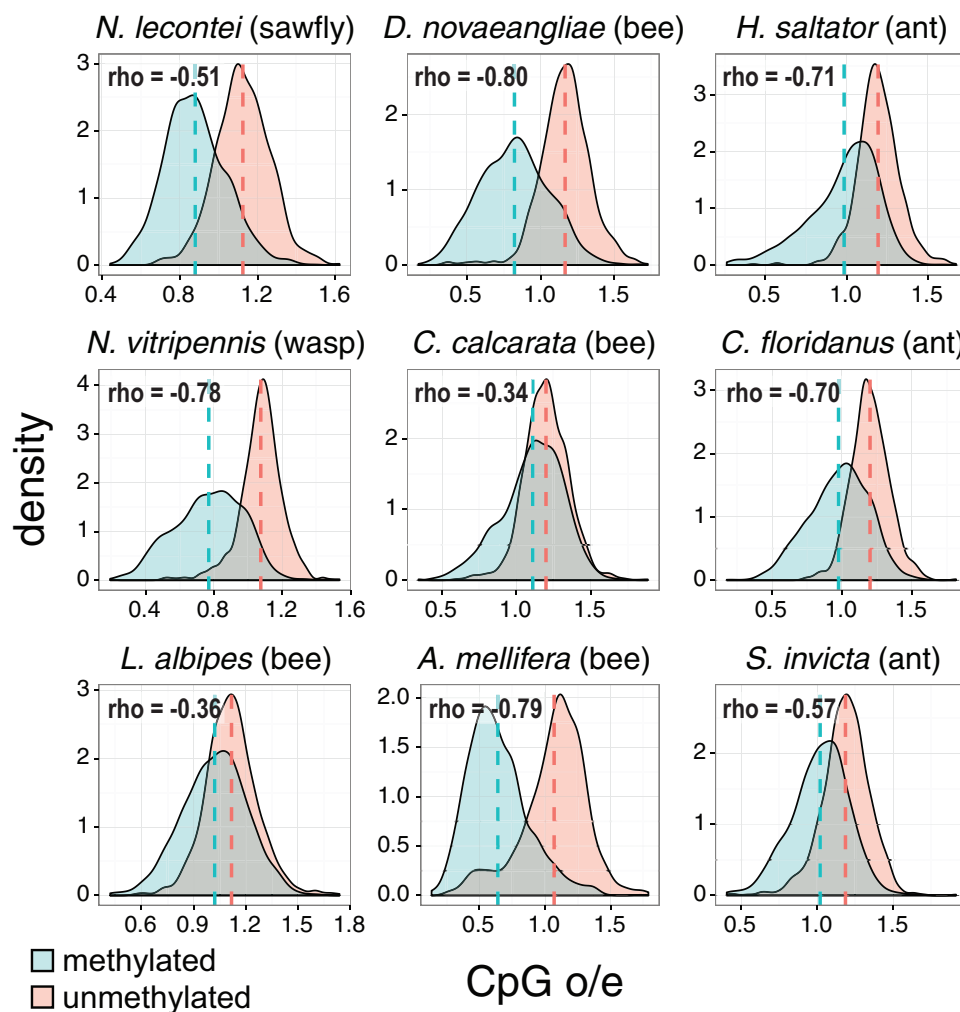


Fig. 1.—Variation in the utility of normalized CpG depletion (CpG o/e) to detect DNA methylation. Density plots of coding sequence CpG o/e measures are shown according to WGBS DNA methylation status ($n = 2400$ ortholog groups with data from all species). Dashed lines represent mean values of CpG o/e and Spearman's rank correlation coefficients (ρ) between CpG o/e and mCG/CG are shown, all significant at $P < 10^{-15}$. For some species, CpG depletion is tightly correlated with DNA methylation, and there is strong separation between methylated and unmethylated genes by CpG o/e values. However, in other species, there is substantial overlap in the CpG o/e values for methylated and unmethylated genes, resulting in a weaker correlation between this metric and DNA methylation.

between coding sequence DNA methylation level and CpG o/e varied extensively, from -0.34 in the bee *C. calcarata* ($P < 10^{-15}$, $n = 2400$ genes) to -0.80 in the bee *Dufourea novaeangliae* ($P < 10^{-15}$, $n = 2400$; fig. 1 and supplementary fig. S2 and table S3, Supplementary Material online). Cluster dendrograms of DNA methylation levels and CpG o/e also exhibited dramatically different topologies (supplementary fig. S5, Supplementary Material online).

One potential explanation for discrepancies in the utility of CpG o/e among taxa is that CpG o/e fails to account for the evolutionary timescale of CpG depletion, and thus may be disproportionately shaped by ancestral patterns of DNA methylation. To better isolate the lineage-specific mutational consequences of DNA methylation, an alternate metric of CpG depletion was calculated as

the transition rate of cytosines to thymines in a CpG context, normalized by the transition rate of cytosines to thymines in a non-CpG context, along terminal branches of our nine species phylogeny. In contrast to CpG o/e measures, this metric is not influenced by CpG depletion that occurred prior to the divergence of sister taxa, and thus should better reflect DNA methylation in extant taxa. Similar to CpG o/e measures, Spearman's rank correlations between this metric and DNA methylation varied substantially, even among sister taxa (e.g., *A. mellifera* $\rho = 0.63$, $P < 10^{-15}$, $n = 958$ and *C. calcarata* $\rho = 0.06$, $P = 0.07$, $n = 958$; supplementary fig. S2, Supplementary Material online). This suggests DNA methylation exhibits variable effects on sequence substitution among taxa over relatively recent evolutionary timescales.

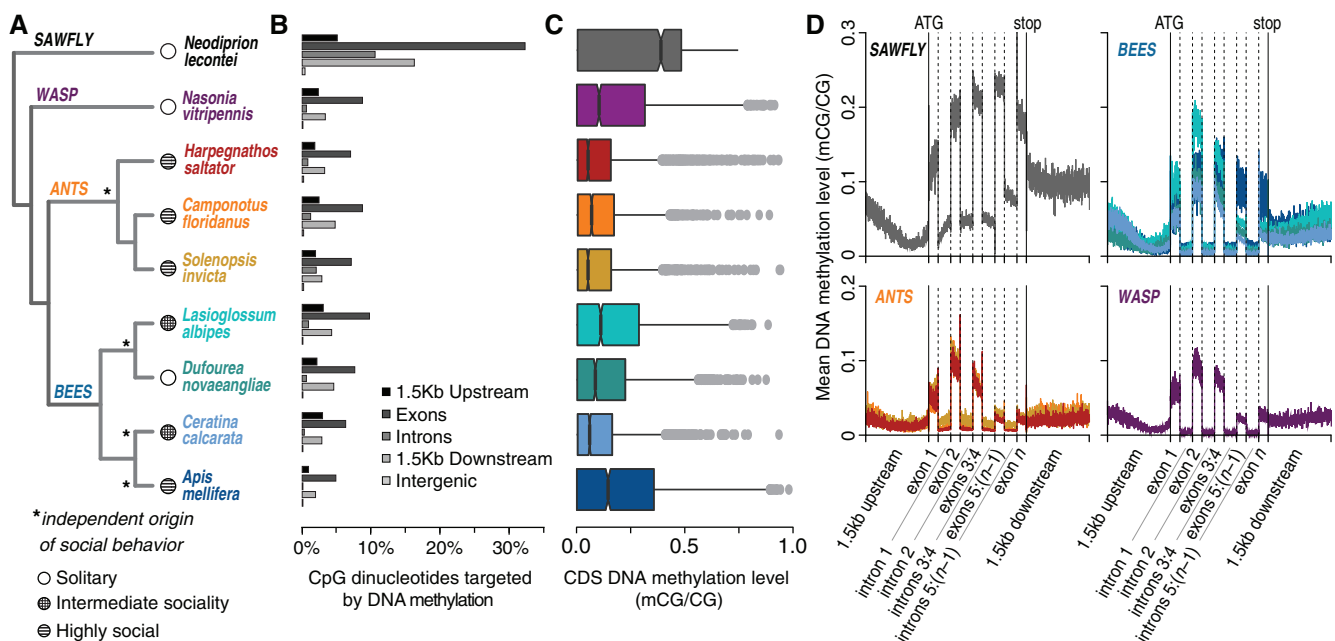


FIG. 2.—Targets and levels of DNA methylation. (A) Cladogram demonstrating species relationships and independent gains of social behavior (social terminology *sensu* Wilson 1971; Kocher and Paxton 2014). (B) Percentage of CpG sites with significant DNA methylation (FDR-corrected binomial test $P < 0.05$) by genomic element classification. (C) Boxplots of coding sequence methylation levels (whiskers show observations within $1.5 \times$ IQR of the lower and upper quartiles; $n = 2400$ ortholog groups with data from all species). (D) Metaplots of mean CpG methylation level over all genes with five or more exons, by position are overlaid for each species.

Among the nine species we investigated, DNA methylation was always most prevalent in exons (fig. 2), consistent with investigations of other holometabolous insects (Xiang et al. 2010; Cunningham et al. 2015). The sawfly *N. lecontei* exhibited surprisingly extensive DNA methylation, with 32% of CpGs within exons targeted by DNA methylation, as compared with 5–10% in the other hymenopterans (fig. 2B; supplementary table S4, Supplementary Material online). *N. lecontei* also exhibited the highest proportion of methylated CpGs in introns, regions upstream and downstream of coding sequences, and intergenic regions (fig. 2B and supplementary table S4, Supplementary Material online).

We also compared DNA methylation levels among taxa when averaged across CpGs in a given coding sequence. DNA methylation levels were significantly higher in *N. lecontei* than in each of the other species (Kruskal–Wallis with Dunn’s Test for Multiple Comparisons $P < 10^{-15}$; fig. 2C and supplementary table S5, Supplementary Material online). In order to test whether differences in DNA methylation on the scale of those detected between species are likely to arise from technical variation or intraspecific biological variation, we assessed previously published WGBS data from multiple castes and tissues of four species in our study. Variation in DNA methylation within a species was not detected on the scale of differences observed between species (supplementary fig. S6, Supplementary Material online). Moreover, *N. lecontei* ranked fourth among taxa in our study in the proportion of genomic

CpG sites with coverage by ten or more reads, illustrating that the pervasiveness of DNA methylation in *N. lecontei* is not an artifact of detection power (supplementary table S6, Supplementary Material online).

Evaluating the Link between DNA Methylation and Social Behavior

Next, we tested whether variation in DNA methylation pervasiveness is associated with reproductive division of labor, as suggested by a recent analysis of CpG o/e variation among ten bee genomes (Kapheim et al. 2015). The taxa we analyzed represent two independent origins of intermediate social behavior (in the halictid bee *L. albipes* [facultatively eusocial] and the xylocopine bee *C. calcarata* [subsocial with facultative alloparental care]), as well as two origins of advanced eusocial behavior (all ants and the honey bee *A. mellifera*; fig. 2A; Materials and Methods). We conducted phylogenetic generalized least squares regression (pGLS) analyses of global and gene-by-gene metrics of DNA methylation versus the degree of sociality exhibited by each taxon (Kocher and Paxton 2014). We found that neither the global mean of CDS methylation nor the proportion of methylated CpGs in any class of genomic element were significantly associated with social complexity ($P > 0.05$ in all cases; table 1). On a gene-by-gene basis, CDS methylation level was also not significantly associated with social complexity for any individual

Table 1

Results of pGLS Analyses of Global DNA Methylation Metrics Versus Level of Sociality Across Hymenopteran Taxa

	R^2	Adjusted R^2	F Value	P Value
CDS global mean mCG/CG	0.1004	-0.1995	0.3347	0.7281
Proportion mCG in 1.5KB upstream of CDS	0.5649	0.4198	3.8940	0.0824
Proportion mCG in exons	0.1541	-0.1279	0.5465	0.6053
Proportion mCG in introns	0.08478	-0.2203	0.2779	0.7666
Proportion mCG in 1.5KB downstream of CDS	0.2143	-0.04761	0.8182	0.4850
Proportion mCG in intergenic regions	0.1657	-0.1124	0.5957	0.5808

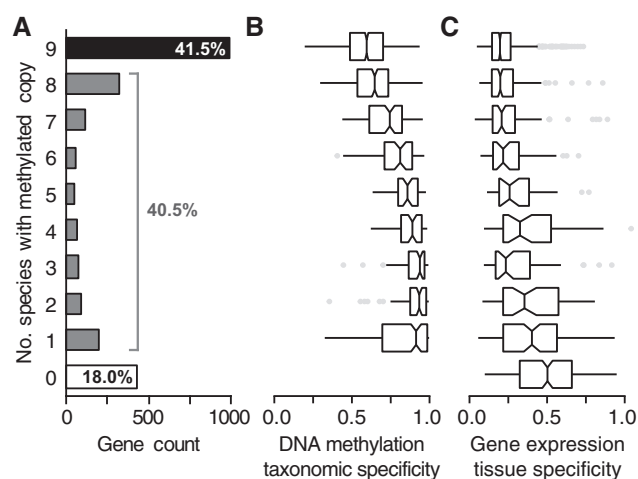


Fig. 3.—Taxonomic specificity of DNA methylation. (A) The number of taxa with DNA methylation in each ortholog group ($n = 2400$). (B) Taxonomic specificity index of DNA methylation level in CDS grouped by the number of taxa with a methylated copy ($n = 1968$ ortholog groups with DNA methylation in at least one taxon; Kruskal–Wallis Rank Sum Test $P < 10^{-15}$). (C) Tissue specificity index of ortholog gene expression among ten tissues in adult *D. melanogaster*, grouped by the number of hymenopteran taxa with a methylated copy ($n = 1676$ *D. melanogaster* orthologs; Kruskal–Wallis Rank Sum Test $P < 10^{-15}$).

ortholog group (FDR $q > 0.1$ in all cases; supplementary fig. S7, Supplementary Material online). These results remained nonsignificant when classifying species simply as “solitary” or “social” (FDR $q > 0.7$ in all cases).

We observed substantial variation among taxa in the genes targeted by DNA methylation. Over 40% of orthologs were targeted by DNA methylation in some, but not all, taxa (fig. 3A). To better quantify the degree of taxonomic variation in DNA methylation, we applied an equation typically used to assess gene expression breadth among tissues, the “tissue specificity index” (Yanai et al. 2005), to DNA methylation levels of orthologous coding sequences across taxa. We termed this metric the “taxonomic specificity index”, which produces values ranging from zero, in the case of uniform DNA methylation levels among taxa, to one, in the case of DNA methylation specific to a single taxon (fig. 3B).

We analyzed ortholog groups belonging to the lowest and highest deciles of DNA methylation taxonomic specificity,

respectively, for enrichment of gene ontology biological process terms. This provided insight into the functions of genes that are stably methylated among species and genes that are variably methylated among species. Genes belonging to ortholog groups with low taxonomic specificity of DNA methylation were enriched for functions associated with translation and biosynthesis (table 2 and supplementary table S7, Supplementary Material online). Genes belonging to ortholog groups with high taxonomic specificity of DNA methylation were enriched for functions associated with cell signaling and behavior (table 2 and supplementary table S7, Supplementary Material online). We further assessed whether taxonomic variation in DNA methylation was associated with the breadth of gene expression among ten adult tissues, as measured in *Drosophila melanogaster* orthologs (Robinson et al. 2013). We observed a significant positive correlation between the gene expression tissue specificity index and the DNA methylation taxonomic specificity index (Spearman’s rank correlation coefficient = 0.24, $P < 10^{-15}$; fig. 3C). To be clear, the genome of *D. melanogaster* lacks CpG methylation (Zemach et al. 2010), so this result does not imply a causal role for DNA methylation in *D. melanogaster* tissue specificity. Instead, this result reveals that DNA methylation is more evolutionarily labile at loci with a narrower expression breadth than is observed for methylated genes as a whole (fig. 3C and supplementary fig. S8, Supplementary Material online). We note that we cannot assume conservation of gene expression breadth between the insect orders Diptera and Hymenoptera. However, we did observe that genes with more specific expression among castes and developmental stages of the ants *C. floridanus* and *H. saltator* also tended to be methylated in fewer taxa than genes with a greater expression breadth (supplementary fig. S9, Supplementary Material online), consistent with our results from *Drosophila* tissue specificity data.

Molecular Evolution of the DNA Methylation Toolkit

The sources of variation in DNA methylation among taxa remain cryptic. One possibility is that variation in DNA methylation patterning is produced by evolution of the DNA methylation or demethylation machinery, namely DNA methyltransferase 1 (DNMT1; Goll and Bestor 2005), DNA

Table 2

Functional Enrichment of Gene Ontology Biological Process Terms for Genes Belonging to the Lowest and Highest Decile of Taxonomic Specificity in Coding Sequence DNA Methylation, When Compared with a Background of All Genes with DNA Methylation in at Least One Taxon

GO Term	Description	log ₁₀ P Value
Lowest decile of DNA methylation taxonomic specificity		
GO:0006412	Translation	−4.5918
GO:0043043	Peptide biosynthetic process	−4.5918
GO:0043603	Cellular amide metabolic process	−3.7595
GO:1901566	Organonitrogen compound biosynthetic process	−3.6289
GO:0044271	Cellular nitrogen compound biosynthetic process	−3.4962
GO:1901576	Organic substance biosynthetic process	−3.0022
Highest decile of DNA methylation taxonomic specificity		
GO:0007186	G-protein coupled receptor signaling pathway	−4.8861
GO:0007411	Axon guidance	−4.8827
GO:0007610	Behavior	−4.5391
GO:0006928	Cellular component movement	−4.4437
GO:0050794	Regulation of cellular process	−3.9666
GO:0007626	Locomotory behavior	−3.5986
GO:0044707	Single-multicellular organism process	−3.5229
GO:0065007	Biological regulation	−3.5114
GO:0045478	Fusome organization	−3.4685
GO:0016310	Phosphorylation	−3.2097
GO:0009611	Response to wounding	−3.1481

Table 3

Results of PAML Branch-Site Tests for Positive Selection on DNMT3, DNMT1, and TET Among Major Clades of Hymenoptera

Gene	Clade	FDR-Corrected P Value	No. BEB Significant Sites
DNMT3	Ants	1.00E+00	0
DNMT3	Bees	1.00E+00	0
DNMT3	Sawflies	5.32E−06*	3
DNMT3	Nonaculeate Wasps	1.00E+00	0
DNMT1	Ants	1.00E+00	28
DNMT1	Bees [a]	1.00E+00	11
DNMT1	Bees [b]	1.00E+00	1
DNMT1	Sawflies	3.44E−02	9
DNMT1	Nonaculeate Wasps	3.44E−02	50
TET	Ants	1.00E+00	11
TET	Bees	1.00E+00	28
TET	Sawflies	1.00E+00	2
TET	Nonaculeate Wasps	1.00E+00	16

NOTE.—DNMT1 experienced a gene duplication event in the bees (Bewick et al. 2017).

*Significant at threshold of FDR-corrected P value <0.01.

methyltransferase 3 (DNMT3; Goll and Bestor 2005), and ten 11 translocation (TET), a 5-methylcytosine oxidase that plays a role in DNA demethylation (Pastor et al. 2013; Wojciechowski et al. 2014). Of particular interest to this question is the de novo methyltransferase DNMT3, which has been shown to play a key role in establishing methylation patterns at CpG sites in animals (Goll and Bestor 2005). Among the DNA methylomes we examined in this study, the sawfly *N. lecontei* is an outlier in two key ways; 1) it exhibits higher levels of DNA methylation and 2) DNA methylation is not biased to the first three exons of genes (fig. 2D). Because the function of DNA

methylation is highly dependent upon genomic context (Jones 2012), such a difference in patterning could reflect increased or distinct functional importance for DNA methylation in *N. lecontei*. Thus, we looked for signatures of enhanced selective constraint or positive selection in DNA methylation machinery of the sawfly clade relative to other hymenopterans (table 3 and supplementary tables S8–S10 and figs. S10–S12, Supplementary Material online).

We performed Phylogenetic Analysis Using Maximum Likelihood (PAML) branch-site tests for positive selection (Zhang et al. 2005; Yang et al. 2005) to compare DNMT1,

DNMT3, and TET evolution among sawfly, ant, bee, and non-acculeate wasp clades. These analyses revealed a strong signature of positive selection only for DNMT3 in the sawfly clade (table 3). This signal was driven by three sites (supplementary table S11, Supplementary Material online). Of particular interest, one site localized to a predicted active site of the ADD domain of DNMT3, which is known to bind unmethylated Histone H3K4, a marker of inactive genes (supplementary table S11, Supplementary Material online; Otani et al. 2009). A second site of interest localized to the DNA methyltransferase domain (supplementary table S11, Supplementary Material online). One or both of these sites could help to explain the differences observed in the methylation levels and patterns observed between *N. lecontei* and other hymenopterans (fig. 2). Because these sites have putatively been subject to positive selection in the sawfly clade as a whole, we hypothesize that patterns of DNA methylation similar to *N. lecontei* will be observed in the other sawfly species in our molecular evolution analysis (*Orussus abietinus* and *Athalia rosae*). Additional taxonomic sampling for DNA methylation analyses will be necessary to test this hypothesis.

Branch-site tests conducted on terminal branches corresponding to species with WGBS data in our study revealed several sites putatively under positive selection in DNMT1 and TET (supplementary tables S9 and S10, Supplementary Material online). Some of these sites localized to the functional domains of these proteins, but none overlapped with predicted active sites (supplementary table S11, Supplementary Material online). No significant positive selection was detected in terminal branches of the DNMT3 tree (supplementary table S8, Supplementary Material online).

PAML branch tests (Yang 2007) identified a significantly lower ratio of the rate of nonsynonymous substitutions to the rate of synonymous substitutions (dN/dS) for the sawfly clade as compared with the tree as a whole for DNMT1, DNMT3, and TET, which suggests there may be enhanced purifying selection operating on DNA methylation machinery in this group (supplementary figs. S10–S12, Supplementary Material online). However, we caution that these tests rely on sequences from substantially fewer species in the sawfly and non-acculeate wasp groups than in ants and bees, resulting in estimates of evolutionary rates from longer branches over longer periods of evolutionary time for these groups.

Discussion

Our study provides new insight into DNA methylation in four major groups of Hymenoptera: sawflies, non-acculeate wasps, bees, and ants. The sawfly *N. lecontei* exhibits by far the highest levels and genomic pervasiveness of DNA methylation discovered in Hymenoptera to date (fig. 2). This highlights exceptional evolutionary variation in DNA methylation, particularly when coupled with the recently documented loss of DNMT3 and dramatic reduction of DNA methylation in

investigated eusocial aculeate wasps (Patalano et al. 2015; Standage et al. 2016). Given that sawflies comprise a sister group to all other hymenopterans, our results suggest that DNA methylation either underwent expansion in the sawfly lineage following the evolutionary divergence of sawflies and other hymenopterans, became depleted in the groups comprising wasps, bees, and ants, or was subject to some combination of these scenarios.

The ortholog groups in our data that exhibit high taxonomic specificity in DNA methylation also exhibit narrower expression breadth among tissues, and are enriched for functions associated with behavior relative to methylated genes as a whole (fig. 3 and table 2). For such ortholog groups, DNA methylation has the potential to contribute to taxon-specific transcriptional regulation and ecologically relevant phenotypes. We note, however, that the relatively narrow expression breadth of genes with evolutionarily labile DNA methylation may be a byproduct of the conservation of DNA methylation at broadly expressed loci (Hunt et al. 2013) rather than indicating an association between variation in DNA methylation and dynamic transcriptional regulation per se.

The mechanisms by which evolutionary variation in DNA methylation arises among hymenopteran taxa remain unclear, but some clues are emerging. The striking preference for DNA methylation targeting to exons (fig. 2D) of constitutively expressed genes suggests commonalities in de novo DNMT3 localization across Hymenoptera (Hunt et al. 2013). As in other eukaryotes (Cedar and Bergman 2009; Otani et al. 2009; Baubec et al. 2015), the patterning of DNA methylation in hymenopteran insects may involve the interaction of DNMT3 functional domains and specific histone modifications (Glastad et al. 2015). Consistent with this idea, our analyses of molecular evolution suggest the relatively distinct DNA methylome of *N. lecontei* may be associated with positive selection on functional DNMT3 domains in the sawfly clade (table 3 and supplementary table S11, Supplementary Material online). As a next step, it will be worthwhile to complement new and diverse hymenopteran DNA methylomes with analyses of DNMT molecular evolution which may ultimately pave the way for functional validation.

A recent study of DNA methylation in two species of *Nasonia* wasps revealed that species-specific patterns of DNA methylation were retained in the parental alleles of F1 hybrids (Wang et al. 2016). This suggests that variation in DNA methylation among closely related species is subject to *cis*-regulation, though this could be limited to the perpetuation of existing patterns by DNMT1 (Wang et al. 2016; Kay et al. 2016). The presence and modality of additional *cis*-regulatory mechanisms, beyond the maintenance of existing marks, could help to explain observed evolutionary variation in DNA methylation and remains an outstanding topic of interest.

Although DNA methylation is known to result in an elevated deamination rate of cytosine to thymine (Shen et al.

1994), the impact of DNA methylation on CpG depletion varies widely between the taxa we investigated (fig. 1 and supplementary figs. S2–S4 and table S3, Supplementary Material online). We find that while there is an overall association between DNA methylation and CpG depletion in all taxa we investigated (supplementary figs. S1 and S2, Supplementary Material online), the strength of this correlation varies greatly among species (fig. 1 and supplementary fig. S2, Supplementary Material online). These results highlight that there are limitations to the utility of CpG o/e, at least when serving as a proxy for DNA methylation in comparative studies of taxa that exhibit sparse DNA methylomes. For example, analyses of CpG o/e alone have suggested that the genome of the facultatively eusocial bee *L. albipes* is nearly (Kapheim et al. 2015) or completely (Bewick et al. 2017) absent of DNA methylation, whereas we find that levels of DNA methylation in *L. albipes* are actually comparable to the genomes of other bees (fig. 2 and supplementary figs. S3 and S4, Supplementary Material online). CpG o/e analyses also overestimated the presence of DNA methylation in the paper wasp *Polistes dominula*, which was subject to ancestral loss of nearly all genomic DNA methylation (Standage et al. 2016).

The sources of variation in CpG depletion resulting from DNA methylation remain unclear, but several factors may contribute. Chief among these is the fact that only mutations arising in germline cells are heritable, and whole body DNA methylation measures may not consistently reflect germline methylation. Variation in effective population size is also expected to influence the efficiency of purifying selection on CpG dinucleotides to maintain amino acid sequence, codon optimality, or a scaffold for DNA methylation. This may be particularly important in insects where the majority of DNA methylation is targeted to coding exons. We note that effective population sizes are thought to be exceptionally low for eusocial insects (Romiguier et al. 2014), which may contribute to increased fixation of nearly neutral mutations through drift. Whatever the explanation for variation among taxa in the effects of DNA methylation on sequence substitution, our results suggest that empirical measures of DNA methylation are essential to confidently assess evolutionary variation in DNA methylation, at least when DNA methylation is present at moderate levels and restricted to coding sequences, as is the case in Hymenoptera.

The results of our investigation are inconsistent with a gross association between DNA methylation and level of sociality among hymenopteran taxa, both at a global level (table 1) and within individual ortholog groups (supplementary fig. S7, Supplementary Material online). A solitary sawfly exhibits more pervasive DNA methylation than other Hymenoptera investigated to date (fig. 2), highly eusocial ants do not exhibit expansion of DNA methylation relative to other Hymenoptera (fig. 2), and some eusocial wasps exhibit greatly reduced DNA methylation and the loss of DNMT3 (Standage et al. 2016;

Patalano et al. 2015). These results may appear initially surprising, given that DNA methylation has been linked to developmental plasticity in some bees and ants (Kucharski et al. 2008; Herb et al. 2012; Alvarado et al. 2015). However, our findings become far less surprising when one considers that a large proportion of methylated genes are broadly expressed among tissues and morphs (Foret et al. 2009; Hunt et al. 2013), and changes in transcriptional regulation at only a few loci may affect developmental outcomes.

DNA methylation need not be an essential precursor to the evolution of social behavior or even exceptionally prevalent in a genome to be coopted on occasion for a role in developmental regulation. Perhaps more importantly, DNA methylation is not unique in its potential to affect developmental gene regulation. DNA methylation, nucleosome positioning, histone protein variants, and histone posttranslational modifications have all been found to influence gene expression by altering the local accessibility of chromatin to transcription factors and the basal transcriptional apparatus (Bintu et al. 2012; Bell et al. 2011). Like DNA methylation, histone modifications play a direct role in the regulation of alternative mRNA splicing (Li-Byarlay et al. 2013; Shukla et al. 2011; Luco et al. 2011). Thus, chromatin states and regulatory outcomes are mediated by a multi-layered and partially redundant epigenetic landscape (Hunt et al. 2013; Maleszka et al. 2014; Glastad et al. 2015; Glastad, Goodisman, et al. 2016), which may help explain the loss of DNA methylation in some insect taxa (Bewick et al. 2017).

There are still many unanswered questions about how DNA methylation and other epigenetic mechanisms evolve. For example, how labile is DNA methylation over evolutionary time? The emerging picture suggests that overall levels of DNA methylation can vary greatly among similar groups of insects, as can the consequences of DNA methylation on genome sequence evolution. DNA methylation appears most evolutionarily labile at loci with a relatively narrow expression breadth and enrichment for behavioral functions. However, the importance of evolutionary variation in DNA methylation to phenotypic diversity in insects remains unknown. Overall, our work helps to create a framework for studying how DNA methylation and other epigenetic factors evolve in the Hymenoptera—a group of insects often cited as a textbook example of epigenetic modification.

Materials and Methods

Samples and DNA Extraction

D. novaeangliae adult females were collected near Lake Ontario in July 2014, *L. albipes* social adult females were collected in Rimont and Aillac, France in August 2013, and *N. lecontei* larval females were collected in Spooner, Wisconsin in July 2014 and reared to adulthood in a laboratory setting. Genomic DNA was extracted from individual whole bodies of

these three species using a standard phenol-chloroform protocol, then pooled for bisulfite conversion and sequencing of a single library per species. *N. lecontei* were subject to egg removal prior to extraction due to their large complement of eggs, which negatively influence DNA quality. Other species sample handling and DNA extraction details can be found in source publications (Wang et al. 2013; Bonasio et al. 2012; Hunt et al. 2013; Zemach et al. 2010; Rehan et al. 2016). We note that the previously published *C. calcarata* data (Rehan et al. 2016) were generated by us from samples that were sequenced at the same time as *D. novaeangliae*, *L. albipes*, and *N. lecontei*. Previously sequenced *S. invicta* data were also generated by us (Hunt et al. 2013). Samples from all species are comprised of adult female bodies (supplementary table S1, Supplementary Material online).

Whole Genome Bisulfite Sequencing (WGBS) and Mapping

Unmethylated enterobacteria phage lambda DNA (GenBank accession: J02459.1) was added to *C. calcarata*, *D. novaeangliae*, *L. albipes*, and *N. lecontei* genomic DNA as a control for bisulfite conversion efficiency. Bisulfite conversion and sequencing for *C. calcarata*, *D. novaeangliae*, *L. albipes*, and *N. lecontei* were performed on the Illumina HiSeq platform by BGI (Shenzhen, China). We used the program Bismark (Krueger and Andrews 2011) to align sequencing reads to each reference genome (Rehan et al. 2016; Kapheim et al. 2015; Kocher et al. 2013; Vertacnik et al. 2016). Reads identified as sequencing duplicates were removed prior to quantification of DNA methylation. For read pairs whose first and second mate overlapped, regions of overlap were counted only once. Read counts were merged between strands, so each CpG was only represented by a single value. For *N. vitripennis*, we used comparable precomputed DNA methylation files (GEO accession GSE43423). New WGBS data generated for this study have been deposited in the Gene Expression Omnibus repository under the accession GSE93893.

DNA Methylation Metrics

Significantly methylated CpG sites were assessed using a binomial test. This test incorporated a bisulfite conversion deamination rate of 0.975 for all species, which conservatively overestimates nonconversion relative to our empirical estimates from *C. calcarata*, *D. novaeangliae*, *L. albipes*, and *N. lecontei* (each of these species had a nonconversion rate of 0.003), as the probability of success. This test assigned a significance value to each CpG site based on the number of unconverted reads (putatively methylated Cs) (Lyko et al. 2010). Resulting *P* values were then adjusted for multiple testing using the method of Benjamini and Hochberg (1995). Only sites with false discovery rate (FDR) corrected binomial *P* values < 0.05 were considered “methylated.” Coding

sequences with three or more methylated sites by this method were considered “methylated” (e.g., fig. 1).

Fractional DNA methylation values were calculated for each CpG dinucleotide as mCG/CG, where mCG is the number of reads with a methylated cytosine at a CpG dinucleotide and CG is the total number of reads mapped to the site. Mean DNA methylation levels were calculated for specific genomic features (e.g., coding sequences) as the mean of all CpG fractional methylation values within that feature (referred to as “methylation level” or “mCG/CG”). For metaplots of mean fractional DNA methylation levels by genic position, exons were divided into 150 proportional bins and introns divided into 200 proportional bins. Mean values within each bin were taken across all genes with five or more exons. Given the documented scarcity of non-CpG methylation in *A. mellifera* (Zemach et al. 2010), *N. vitripennis* (Wang et al. 2013), *H. saltator* (Bonasio et al. 2012), and *C. floridanus* (Bonasio et al. 2012), we restricted our analyses of DNA methylation to CpG dinucleotides.

CpG Depletion

Normalized CpG depletion (CpG o/e) was calculated for genomic elements as $P_{\text{CpG}}/(P_{\text{C}} * P_{\text{G}})$, where P_{CpG} , P_{C} , and P_{G} are the frequencies of CpG, cytosine, and guanine, respectively (Elango et al. 2009; Yi and Goodisman 2009). Evolutionary transitions from C to T were assessed as follows. Sequence alignments were used to reconstruct the ancestral state at each nucleotide for each node in the nine species phylogenetic tree using PRANK’s ancestral state reconstruction (Löytynoja and Goldman 2005). We assessed nucleotide changes between a given species and its closest ancestral sequence using a custom script. The transition rate of cytosines in a CpG context normalized by the transition rate of cytosines in a non-CpG context, along terminal branches of our nine species phylogeny was assessed as the metric “proportion of CpG → T/proportion of CpH → T.”

Ortholog Assignment

We used a reciprocal BLAST approach to assign sequences in *C. calcarata* and *N. lecontei* with homology to putative 1-to-1 ortholog groups established among the other hymenopterans by OrthoDB (Kriventseva et al. 2015). *D. melanogaster* orthologs were also determined by OrthoDB. In the case of multiple *D. melanogaster* orthologs mapping to an ortholog group, one *D. melanogaster* ortholog was taken at random.

DNA Methylation and Level of Sociality

We categorized species as “solitary”, “intermediately social”, or “highly social” based on criteria related to reproductive division of labor (Wilson 1971; Kocher and Paxton 2014). The sawfly *N. lecontei*, the wasp *N. vitripennis*, and the bee *D. novaeangliae* are each absent reproductive division of labor

and were classified as solitary by these criteria. The bee *C. calcarata* is not eusocial, but exhibits prolonged maternal care and facultative alloparental care (Rehan et al. 2014), and the bee *L. albipes* is facultatively eusocial (Kocher et al. 2013). Both *L. albipes* and *C. calcarata* were classified as intermediately social. The honey bee *A. mellifera* and the ants *H. saltator*, *C. floridanus*, and *S. invicta* all exhibit advanced eusociality marked by the obligate presence of a nonreproductive worker caste, and were classified as highly social.

To assess relationships between DNA methylation and the three levels of sociality described above while accounting for underlying phylogenetic correlations, we employed a phylogenetic generalized least squares regression (pGLS) analysis using the R package, *caper* (Orme 2013). Lambda was estimated concurrently via maximum-likelihood. Raw *P* values were corrected for multiple testing using the method of Benjamini and Hochberg (1995). Alternate analyses were also conducted in which the “intermediate sociality” and “highly social” levels were joined simply as “social” and produced similar results.

DNA Methylation and Gene Expression Specificity Indices

The gene expression tissue specificity index (Yanai et al. 2005) for *D. melanogaster* was calculated among 10 adult tissues in FlyAtlas (Robinson et al. 2013), as:

$$\frac{\sum_{j=1}^n 1 - (E_j/E_{\max})}{n - 1},$$

where *n* is the number of tissues, E_j is the expression level of the gene in the *j*th tissue and E_{\max} the maximum expression level of the gene across the *n* tissues. A gene expression “caste and developmental stage” specificity index was also calculated in the same manner for the ants *C. floridanus* and *H. saltator*, for those genes with total FPKM > 1. *C. floridanus* and *H. saltator* RNA-seq reads were previously published (Bonasio et al. 2010) and FPKM values were generated as described in Glastad et al. (2015). A combination of five adult sample-types and two developmental stages were assessed for *C. floridanus*, and a combination of four adult sample-types and two developmental stages were assessed for *H. saltator* (supplementary fig. S9, Supplementary Material online).

The DNA methylation taxonomic specificity index among species was calculated as:

$$\frac{\sum_{j=1}^n 1 - [(mCG/CG)_j / (mCG/CG)_{\max}]}{n - 1},$$

where *n* is the number of species, $(mCG/CG)_j$ is the CDS DNA methylation level of the gene in the *j*th species and

$(mCG/CG)_{\max}$ is the maximum CDS DNA methylation level of the gene across the *n* species.

Functional Enrichment

Gene Ontology (GO) biological process functional enrichment analysis was performed using *D. melanogaster* ortholog gene identifiers to create target and background lists. Gene lists were analyzed for enrichment of GO terms with the GOrilla tool (Eden et al. 2009). Full lists of enriched GO terms (supplementary table S7, Supplementary Material online) were subsequently filtered to remove redundant terms using ReviGO (table 2; Supek et al. 2011).

Global Phylogenetics and Coding Sequence Evolution

A phylogeny of our target species was generated from fourfold degenerate sites of all shared 1:1 orthologs across all species. Coding sequences were aligned using PRANK (Löytynoja and Goldman 2005). Alignments were then run through GBLOCKS ($-t=c$ $-b4=6$ $-b5=h$; Talavera and Castresana 2007) to filter low quality alignments. Fourfold degenerate sites were isolated from each ortholog and concatenated. These were then used for tree estimation in RAxML (Stamatakis 2014) using a mixed/partition model with GTRGAMMA model for bootstrapping with autoMRE. The majority-rule consensus tree (supplementary fig. S5A, Supplementary Material online) was used for all downstream analyses.

GBLOCKS-filtered alignments were then input into RAxML for gene tree construction. The resulting topologies were compared with the species phylogeny using the comparison tools in FastTree (Price et al. 2009). Only alignments with topologies matching the species tree were included in downstream PAML analyses. PAML (Yang 2007) was used to estimate synonymous and nonsynonymous substitution rates (dS and dN) with free ratios for each branch. Terminal branch values were used to represent the evolutionary rates of coding sequences in each species for alignments longer than 50 bases (supplementary figs. S2 and S8, Supplementary Material online).

DNMT 1, DNMT3, and TET Molecular Evolution

Gene sequences for DNMT1, DNMT3, and TET were identified by a homology search using BLASTp to query known sequences from a bee, *A. mellifera* (DNMT1, DNMT3, and TET), a beetle, *Nicrophorus vespilloides* (DNMT1, DNMT3), and a fly, *Drosophila melanogaster* (TET) against target gene sets from 31 hymenopteran taxa (supplementary figs. S10–S12, Supplementary Material online). Prospective proteins were then run through InterProScan (Jones et al. 2014) to ensure that characteristic protein domains were present. Multiple sequence alignments were performed on the amino acid sequence using PASTA (Mirarab et al. 2015) and were

then back-translated to codon sequences. Conserved blocks were extracted using Gblocks (Talavera and Castresana 2007). RAxML was used with the GTRGAMMA model, to estimate a gene tree for each gene (Stamatakis 2014). PAML (Yang 2007) branch tests were performed on the ant, bee (both bee clades for DNMT1), wasp, and sawfly clades, as well as the species with WGBS data in our study (null: model = 0, NSsites = 0, fix_omega = 0, omega = 1; alternative: model = 2, NSsites = 0, fix_omega = 0, omega = 1; supplementary figs. S10–S12, Supplementary Material online). The branch-site test A for positive selection (Zhang et al. 2005) was performed on the same branches and clades (null: model = 2, NSsites = 2, fix_omega = 1, omega = 1; alternative: model = 2, NSsites = 2, fix_omega = 0, omega = 1; table 2 and supplementary tables S8–S11, Supplementary Material online).

Supplementary Material

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

Author Contributions

K.M.G., S.D.K., and B.G.H. designed the study. K.M.G., S.D.K., B.G.H., and S.V.A. analyzed the data. All authors contributed resources. S.D.K. and B.G.H. wrote the paper with contributions from K.M.G. and S.V.A. and feedback from all authors.

Acknowledgments

This work was supported by the University of Georgia College of Agricultural and Environmental Sciences and Princeton University. We thank Mary Centrella, Achik Dorchin, and Graham Montgomery for *Dufourea novaeangliae* collection, John Terbot for *Neodiprion lecontei* collection, Adam Bewick for assistance with molecular evolution analysis, and the Georgia Advanced Computing Resource Center for computational resources and support.

Literature Cited

- Alvarado S, Rajakumar R, Abouheif E, Szyf M. 2015. Epigenetic variation in the *Egfr* gene generates quantitative variation in a complex trait in ants. *Nat Commun.* 6:6513.
- Baubec T, et al. 2015. Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* 520:243–247.
- Bell O, Tiwari VK, Thomä NH, Schubeler D. 2011. Determinants and dynamics of genome accessibility. *Nat Rev Genet.* 12:554–564.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc B* 57:289–300.
- Berger SL, Kouzarides T, Shiekhattar R, Shilatifard A. 2009. An operational definition of epigenetics. *Genes Dev.* 23:781–783.
- Bewick AJ, Vogel KJ, Moore AJ, Schmitz RJ. 2017. Evolution of DNA methylation across insects. *Mol Biol Evol.* 34:654–665.
- Bintu L, et al. 2012. Nucleosomal elements that control the topography of the barrier to transcription. *Cell* 151:738–749.
- Bonasio R, et al. 2012. Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Curr Biol.* 22:1755–1764.
- Bonasio R, et al. 2010. Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Science* 329:1068–1071.
- Cedar H, Bergman Y. 2009. Linking DNA methylation and histone modification: patterns and paradigms. *Nat Rev Genet.* 10:295–304.
- Cunningham CB, et al. 2015. The genome and methylome of a beetle with complex social behavior, *Nicrophorus vespilloides* (Coleoptera: Silphidae). *Genome Biol Evol.* 7:3383–3396.
- Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10:48.
- Elango N, Hunt BG, Goodisman MAD, 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.* 106:11206–11211.
- Foret S, Kucharski R, Pittelkow Y, Lockett GA, Maleszka R. 2009. Epigenetic regulation of the honey bee transcriptome: unravelling the nature of methylated genes. *BMC Genomics* 10:472.
- Gadau J, et al. 2012. The genomic impact of 100 million years of social evolution in seven ant species. *Trends Genet.* 28:14–21.
- Glastad KM, Gokhale K, Liebig J, Goodisman MAD. 2016. The caste- and sex-specific DNA methylome of the termite *Zootermopsis nevadensis*. *Sci Rep.* 6:37110.
- Glastad KM, Goodisman MAD, Yi SV, Hunt BG. 2016. Effects of DNA methylation and chromatin state on rates of molecular evolution in insects. *G3* 6:357–363.
- Glastad KM, Hunt BG, Goodisman MAD. 2015. DNA methylation and chromatin organization in insects: insights from the ant *Camponotus floridanus*. *Genome Biol Evol.* 7:931–942.
- Glastad KM, Hunt BG, Yi SV, Goodisman MAD. 2011. DNA methylation in insects: on the brink of the epigenomic era. *Insect Mol Biol.* 20:553–565.
- Goll MG, Bestor TH. 2005. Eukaryotic cytosine methyltransferases. *Annu Rev Biochem.* 74:481–514.
- Herb BR, et al. 2012. Reversible switching between epigenetic states in honeybee behavioral subcastes. *Nat Neurosci.* 15:1371–1373.
- Hunt BG, Glastad KM, Yi SV, Goodisman MAD. 2013. Patterning and regulatory associations of DNA methylation are mirrored by histone modifications in insects. *Genome Biol Evol.* 5:591–598.
- Jaenisch R, Bird A. 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet.* 33:245–254.
- Jones P, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30:1236–1240.
- Jones PA. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet.* 13:484–492.
- Kapheim KM, et al. 2015. Genomic signatures of evolutionary transitions from solitary to group living. *Science* 348:1139–1143.
- Kay S, Skowronski D, Hunt BG. 2016. Developmental DNA methyltransferase expression in the fire ant *Solenopsis invicta*. *Insect Sci.* doi: 10.1111/1744-7917.12413.
- Kocher SD, et al. 2013. The draft genome of a socially polymorphic halictid bee, *Lasioglossum albipes*. *Genome Biol.* 14:R142.
- Kocher SD, Paxton RJ. 2014. Comparative methods offer powerful insights into social evolution in bees. *Apidologie* 45:289–305.
- Kriventseva EV, et al. 2015. OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. *Nucleic Acids Res.* 43:D250–D256.
- Krueger F, Andrews SR. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27:1571–1572.

- Kucharski R, Maleszka J, Foret S, Maleszka R. 2008. Nutritional control of reproductive status in honeybees via DNA methylation. *Science* 319:1827–1830.
- 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*. 110:12750–12755.
- Libbrecht R, Oxley PR, Keller L, Kronauer DJC. 2016. Robust DNA methylation in the clonal raider ant brain. *Curr Biol*. 26:391–395.
- Löytynoja A, Goldman N. 2005. An algorithm for progressive multiple alignment of sequences with insertions. *Proc Natl Acad Sci U S A*. 102:10557–10562.
- Luco RF, Allo M, Schor IE, Kornblihtt AR, Misteli T. 2011. Epigenetics in alternative pre-mRNA splicing. *Cell* 144:16–26.
- Lyko F, et al. 2010. The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS Biol*. 8:e1000506.
- Maleszka R, Mason PH, Barron AB. 2014. Epigenomics and the concept of degeneracy in biological systems. *Brief Funct Genomics* 13:191–202.
- Mirarab S, et al. 2015. PASTA: ultra-large multiple sequence alignment for nucleotide and amino-acid sequences. *J Comput Biol*. 22:377–386.
- Orme D. 2013. The caper package: comparative analysis of phylogenetics and evolution in R. Available from: <https://cran.r-project.org/web/packages/caper/index.html>.
- Otani J, et al. 2009. Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. *EMBO Rep*. 10:1235–1241.
- Pastor WA, Aravind L, Rao A. 2013. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat Rev Mol Cell Biol*. 14:341–356.
- Patalano S, et al. 2015. Molecular signatures of plastic phenotypes in two eusocial insect species with simple societies. *Proc Natl Acad Sci U S A*. 112:13970–13975.
- Price MN, Dehal PS, Arkin AP. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol*. 26:1641–1650.
- Rehan SM, Glastad KM, Lawson SP, Hunt BG. 2016. The genome and methylome of a subsocial small carpenter bee, *Ceratina calcarata*. *Genome Biol. Evol*. 8:1401–1410.
- Rehan SM, Berens AJ, Toth AL. 2014. At the brink of eusociality: transcriptomic correlates of worker behaviour in a small carpenter bee. *BMC Evol Biol*. 14:260.
- Robinson SW, Herzyk P, Dow JAT, Leader DP. 2013. FlyAtlas: database of gene expression in the tissues of *Drosophila melanogaster*. *Nucleic Acids Res*. 41:D744–D750.
- Romiguier J, et al. 2014. Population genomics of eusocial insects: the costs of a vertebrate-like effective population size. *J Evol Biol*. 27:593–603.
- Shen JC, Rideout WM, Jones PA. 1994. The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucleic Acids Res*. 22:972–976.
- Shukla S, et al. 2011. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* 479:74–79.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.
- 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*. 25:1769–1784.
- Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6:e21800.
- Suzuki MM, Bird A. 2008. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet*. 9:465–476.
- Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol*. 56:564–577.
- Vertacnik KL, Geib SM, Linnen CR. 2016. *Neodiprion lecontei* genome assembly Nlec1.0. NCBI/GenBank.
- Wang H, et al. 2012. Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res*. 22:1680–1688.
- Wang X, et al. 2013. Function and evolution of DNA methylation in *Nasonia vitripennis*. *PLoS Genet*. 9:e1003872.
- 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:e1002500.
- Wilson EO. 1971. *The Insect Societies*. The Belknap Press of Harvard University Press: Cambridge, MA.
- Wojciechowski M, et al. 2014. Insights into DNA hydroxymethylation in the honeybee from in-depth analyses of TET dioxygenase. *Open Biol*. 4:140110.
- Xiang H, et al. 2010. Single base-resolution methylome of the silkworm reveals a sparse epigenomic map. *Nat Biotechnol*. 28:516–520.
- Yanai I, et al. 2005. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* 21:650–659.
- Yang Z, Wong WSW, Nielsen R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol Biol Evol*. 22:1107–1118.
- Yang ZH. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24:1586–1591.
- Yi SV, Goodisman MAD. 2009. Computational approaches for understanding the evolution of DNA methylation in animals. *Epigenetics* 4:551–556.
- Zemach A, McDaniel IE, Silva P, Zilberman D. 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328:916–919.
- Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol*. 22:2472–2479.
- Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S. 2007. Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet*. 39:61–69.

Associate editor: Dan Graur