Functional Conservation of DNA Methylation in the Pea Aphid and the Honeybee

Brendan G. Hunt¹, Jennifer A. Brisson², Soojin V. Yi*,¹, and Michael A. D. Goodisman*,¹

¹School of Biology, Georgia Institute of Technology
²School of Biological Sciences, University of Nebraska

*Corresponding author: E-mail: michael.goodisman@biology.gatech.edu; soojin.yi@biology.gatech.edu.

Accepted: 13 September 2010

Abstract

DNA methylation is a fundamental epigenetic mark known to have wide-ranging effects on gene regulation in a variety of animal taxa. Comparative genomic analyses can help elucidate the function of DNA methylation by identifying conserved features of methylated genes and other genomic regions. In this study, we used computational approaches to distinguish genes marked by heavy methylation from those marked by little or no methylation in the pea aphid, Acyrthosiphon pisum. We investigated if these two classes had distinct evolutionary histories and functional roles by conducting comparative analysis with the honeybee, Apis (Ap.) mellifera. We found that highly methylated orthologs in A. pisum and Ap. mellifera exhibited greater conservation of methylation status, suggesting that highly methylated genes in ancestral species may remain highly methylated over time. We also found that methylated genes tended to show different rates of evolution than unmethylated genes. In addition, genes targeted by methylation were enriched for particular biological processes that differed from those in relatively unmethylated genes. Finally, methylated genes were preferentially ubiquitously expressed among alternate phenotypes in both species, whereas genes lacking signatures of methylation were preferentially associated with condition-specific gene expression. Overall, our analyses support a conserved role for DNA methylation in insects with comparable methylation systems.

Key words: comparative genomics, DNA methylation, epigenetics, insects, phenotypic plasticity, polyphenism.

Introduction

DNA methylation is an important epigenetic modification that plays a role in gene regulation in many organisms (Wolffe and Matzke 1999; Jaenisch and Bird 2003; Weber et al. 2007). Although DNA methylation occurs in all three domains of life, its genomic patterns show considerable variation among taxa (Hendrich and Tweedie 2003; Field et al. 2004; Suzuki and Bird 2008). For example, vertebrate genomes exhibit global patterns of methylation, but invertebrate genomes tend to display reduced or minimal levels of methylation (Suzuki and Bird 2008). Moreover, methylation of gene promoter regions in vertebrates leads to transcriptional repression (Wolffe and Matzke 1999; Jaenisch and Bird 2003; Weber et al. 2007; Zemach et al. 2010), but this relationship has not been observed in invertebrates. Instead, methylation primarily targets invertebrate gene bodies (Suzuki and Bird 2008; Xiang et al. 2010; Zemach et al. 2010). These contrasting patterns and effects have traditionally enforced the view that DNA methylation plays a fundamentally different role in vertebrate and invertebrate genomes.

The arrival of genome sequences from multiple insects now makes a greater understanding of the patterns and phenotypic consequences of DNA methylation more tangible (Honeybee Genome Sequencing Consortium 2006; Wang et al. 2006; The International Aphid Genomics Consortium 2010; The Nasonia Genome Working Group 2010; Walsh et al. 2010). Specifically, comparative genomic analysis can be used to determine whether targets of DNA methylation are conserved between taxa. Moreover, the inferred patterns of methylation can be used to test current hypotheses explaining the evolutionary persistence of DNA methylation (Yi and Goodisman 2009). For example, it has been hypothesized that gene body methylation may act to minimize spurious transcription patterns (Suzuki et al. 2007; Maunakea et al. 2010), which could explain observations of dense methylation in functionally conserved genes and genes with ubiquitous expression among tissues.
in invertebrates (Suzuki et al. 2007; Foret et al. 2009; Xiang et al. 2010). It has also been suggested that DNA methylation persists in animals for genomic defense against transposable elements (Yoder et al. 1997, but see Regev et al. 1998; Simmen et al. 1999; Suzuki et al. 2007, and Xiang et al. 2010). DNA methylation may also act as an important mechanism for genomic imprinting, which results in the differential expression of parental alleles (Reik and Walter 2001). Finally, de novo DNA methylation is hypothesized to play an important role in developmental responsiveness to environmental factors and the regulation of phenotypic plasticity, as is apparently the case in the honeybee (Jaenisch and Bird 2003; Kucharski et al. 2008; Maleszka 2008).

The purpose of this study was to determine whether DNA methylation plays a conserved role in divergent insects with comparable DNA methylation systems. We provided insight into this question by comparing and contrasting the evolutionary signatures of DNA methylation in the genomes of the pea aphid, *Acyrthosiphon pisum*, and the honeybee, *Apis* (*A.* mellifera).

*Acyrthosiphon pisum* diverged from *Ap. mellifera* more than 300 Ma (Gaunt and Miles 2002; Honeybee Genome Sequencing Consortium 2006), a time frame roughly equivalent to the divergence of modern birds and mammals (Kumar and Hedges 1998). Developmentally, *Ap. mellifera* undergoes full metamorphosis and possesses morphologically distinct larval, pupal, and adult stages. In contrast, *A. pisum* develops gradually and does not undergo metamorphosis. However, *A. pisum* and *Ap. mellifera* both serve as important models for understanding the evolution and development of phenotypic plasticity (Evans and Wheeler 2001; Brisson and Stern 2006; Honeybee Genome Sequencing Consortium 2006; Brisson 2010; The International Aphid Genomics Consortium 2010).

Specifically, aphids have a complex life cycle that alternates between asexual and sexual development. Asexual females exhibit a wing polyphenism in which they produce either winged or unwinged morphs depending on environmental cues (reviewed in Müller et al. 2001). During the sexual portion of the life cycle, males also produce winged or unwinged morphs. However, morph determination is genetic in males, and thus male wing dimorphism is referred to as a polymorphism (Smith and MacKay 1989). Honeybees, on the other hand, are highly social and dwell in large, predominantly female, colonies (Wilson 1971). Individuals partake in a remarkable division of labor, with a single queen typically dominating reproduction and workers engaged in tasks related to brood rearing, foraging, and colony defense (Wilson 1971). Queen and worker castes are developmentally determined by nutritional factors and exhibit dramatically different anatomy and behavior (Wheeler 1986; Evans and Wheeler 2001).

Importantly, both *Ap. mellifera* and *A. pisum* show evidence of widespread DNA methylation that is predominantly targeted to genes (Wang et al. 2006; Elango et al. 2009; Walsh et al. 2010). Consequently, patterns of genome methylation in *A. pisum* and *Ap. mellifera* can provide considerable insight into the function of gene methylation in insects, in particular, and invertebrates, in general.

In this study, we investigated the conservation of DNA methylation patterns in *A. pisum* and *Ap. mellifera* by first testing whether genes with similar functions are targeted by DNA methylation in both species. To achieve this aim, we examined patterns of functional enrichment among genes marked by relatively dense methylation and relatively sparse methylation. We further tested whether shared patterns of functional enrichment among DNA methylation targets are associated with conservation at the sequence level (Suzuki et al. 2007). Next, we examined whether *A. pisum* provided support for the hypothesis that genes with sparse methylation exhibit condition-specific gene expression (Elango et al. 2009; Foret et al. 2009). Finally, we synthesized our results with those from other recent investigations to advance a more comprehensive understanding of DNA methylation in insects. Overall, our results provide support for a remarkable level of conservation in gene methylation status and function over evolutionary time.

### Materials and Methods

#### Gene Sequences

Analyses were conducted on mRNA transcript sequences because evidence suggests that DNA methylation preferentially targets exons in insects and other invertebrates (Wang et al. 2006; Suzuki et al. 2007; Elango et al. 2009; Xiang et al. 2010; Zemach et al. 2010). For *A. pisum*, the “ACYPmRNA” and the “ACYPproteins” official genes consensus sets were obtained from AphidBase (http://www.aphidbase.com). For *Ap. mellifera*, the “Amel_pre_release2” amino acid sequence official gene set (OGS) was obtained from BeeBase (http://www.beebase.org), and model RefSeq transcripts were downloaded from the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/Ftp). *Apis mellifera* OGS IDs were converted to RefSeq accessions using the “gene_info” and “gene2refseq” databases, also available from NCBI. For *Drosophila melanogaster*, “Release_5.21” transcript and protein sequence sets were obtained from flybase (http://flybase.org).

#### Normalized CpG Dinucleotide Content (CpGOE)

We used CpGOE as a measure of the level of DNA methylation of genes (Saxonov et al. 2006; Suzuki et al. 2007; Weber et al. 2007; Yi and Goodisman 2009). CpGOE acts
as a metric of levels of DNA methylation because methylation occurs predominantly on CpG dinucleotides in animals and methylated cytosines are hypermutable due to spontaneous deamination. This deamination causes a gradual depletion of CpG dinucleotides from methylated regions over time (Bird 1980). Consequently, genomic regions with relatively dense germline methylation have low CpGDE and regions with little or no germline methylation maintain high levels of CpGDE. It is important to note that CpGDE could be influenced by either the number of methylated CpG sites or the proportion of cells incurring methylation at a given locus. In addition, somatic mutations are not transmitted to progeny and therefore cannot influence CpGDE in and of themselves. However, CpGDE has been linked to empirically determined levels of DNA methylation in somatic tissues in insects, suggesting that many genes are universally methylated in germlines and soma (Foret et al. 2009; Xiang et al. 2010).

CpGDE was calculated as described previously (Elango et al. 2009), from the gene sets above. Only RefSeq model sequences were used for analyses involving CpGDE in A. pisum (except in the case of gene expression analysis, described below) because RefSeq models were used for Ap. mellifera in our analysis. Sequences with CpGDE values of 0 were removed from further analysis.

Bimodal distributions of CpGDE have previously been reported in both Ap. mellifera (Elango et al. 2009; Foret et al. 2009; Wang and Leung 2009) and A. pisum (Walsh et al. 2010). In this study, we used the NOCOM software package (Ott 1979) to estimate means, standard deviations, and proportions of two components of the mixture of normal distributions of CpGDE for both A. pisum and Ap. mellifera. These distributions were plotted using R (R Development Core Team 2010), and their intersections were used as cutoffs to divide low CpGDE and high CpGDE gene classes.

**Orthology**

Three-way orthologs between A. pisum, Ap. mellifera, and D. melanogaster were identified by first performing pairwise BlastP comparisons of complete protein sequence sets with a cutoff of $1 \times 10^{-5}$, next identifying pairwise reciprocal best hits, and finally identifying orthologs with shared best hits among all pairwise comparisons (Altschul et al. 1997; Stajich et al. 2002). Orthologs determined in this manner were used for comparisons of CpGDE and evolutionary distance between orthologs from A. pisum and Ap. mellifera.

Pairwise orthologs shared between A. pisum and D. melanogaster were identified by performing BlastP comparisons of complete protein sequence sets with a cutoff of $1 \times 10^{-5}$ and identifying reciprocal best hits. Only orthologs with RefSeq model proteins in A. pisum were retained.

**Sequence Divergence**

In order to compare the evolutionary divergence of low CpGDE and high CpGDE orthologs between A. pisum and Ap. mellifera, a total of 2,222 orthologous protein sequences were first aligned using ClustalW (Thompson et al. 1994). Confidently, aligned gap-free columns were then extracted using Gblocks with default settings (Castresana 2000), and only long alignments ($\geq 100$ amino acids) were kept for analysis. PAL2NAL was used to convert protein sequence alignments to corresponding codon alignments (Suyama et al. 2006). Finally, PAML was used to calculate rates of synonymous (dS) and nonsynonymous (dN) substitution with the “codeml” method (Yang 2007). Because synonymous substitution rates were predominantly saturated (dS $> 2$), measures of dN and DNA sequence percent identity were used to assess sequence divergence.

**Gene Ontology**

Gene ontology (GO) annotations for D. melanogaster orthologs of A. pisum proteins were used to analyze enrichment of biological process terms (Ashburner et al. 2000). GO biological process term enrichment was determined by comparing orthologs of low CpGDE and high CpGDE genes separately with a background composed of both low CpGDE and high CpGDE orthologs using the DAVID bioinformatics database functional annotation tool (Dennis et al. 2003). A Benjamini multiple-testing correction of the EASE score (a modified Fisher exact P value; Hosack et al. 2003) was used to determine statistical significance of GO term enrichment.

**EST Mapping**

Acyrthosiphon pisum expressed sequence tags (ESTs), previously used to characterize differential gene expression underlying developmental differences, sex differences, female wing polyphenism, and wing morph differences (Brisson et al. 2007), were mapped to the A. pisum official genes consensus set (OGS) to aid in assessing the relationship between the degree of differential gene expression among phenotypic classes and CpGDE. EST sequences were compared with all OGS mRNA sequences by BlastN (Altschul et al. 1997). To be considered a match, EST query sequences were required to have $>50\%$ sequence alignment to an OGS hit, $>95\%$ identity of the aligned sequence, and reciprocal best hits resulting from BlastN analysis of the OGS query against an EST database. GLEAN as well as RefSeq gene models were accepted in this case to map a greater proportion of microarray data.

**Gene Expression**

Brisson et al. (2007) previously examined the gene expression differences underlying distinct phenotypes in A. pisum
using cDNA microarrays (Wilson et al. 2006). Specifically, microarrays were utilized to determine the degree of differential gene expression in comparisons of 1) fourth instar juveniles versus adults (compared within unwinged males, within winged males, within unwinged asexual females, and within winged asexual females), 2) males versus asexual females (compared within winged fourth instars, within unwinged fourth instars, within winged adults, and within unwinged adults), 3) polyphenic winged versus unwinged females (compared within fourth instars and within adults), and finally, polymorphic winged versus unwinged males (compared within fourth instars and within adults).

For the present study, we calculated the mean of the absolute value of log₂-transformed ratios across multiple comparisons to measure the degree of differential gene expression. In this manner, we combined data from all pairwise comparisons of 1) development, 2) sex, 3) female wing polymorphism, and 4) male wing polymorphism. The mean of log₂-transformed gene expression ratios across all 12 pairwise comparisons was also calculated. We further divided each of these measures into two bins at a mean [log₂ expression ratio] value of 0.5, with genes below this threshold roughly corresponding to genes with similar expression between groups and genes above this value roughly corresponding to genes with differential expression between groups.

We also revisited analysis previously described and published by Elango et al. (2009), which demonstrated that high CpG_E genes were overrepresented among genes that were differentially expressed between queen and worker castes (Grozinger et al. 2007). For the present manuscript, we analyzed NCBI transcript sequences rather than introns and exons combined, to remain consistent with our analyses of aphid gene expression.

Finally, Foret et al. (2009) previously used an oligonucleotide microarray representing the honeybee OGS (Honeybee Genome Sequencing Consortium 2006) to assess the expression breadth of genes among the following tissues in Ap. mellifera: antenna, brain, whole-body larva, hypopharyngeal gland, ovary, and thorax. They further demonstrated that low CpG_E genes were vastly overrepresented among genes with ubiquitous expression (Foret et al. 2009). We expanded upon their analysis by splitting genes into six classes based upon the number of tissues with observed expression. To do so, we utilized lists of genes expressed in each tissue, along with a fasta file of sequences used to design the array. To map sequences with generic microarray identifiers to honeybee model RefSeq transcripts, we compared the sequences using BlastN (Altschul et al. 1997). To be considered a match, array query sequences were required to have >50% sequence alignment to a model RefSeq transcript hit and >98% identity for the aligned sequence. We then generated a numeric count of the number of tissues in which each gene was expressed (integers from 1 to 6) and recorded the CpG_OE for each associated model RefSeq transcript. Data for expression breadth and CpG_OE were obtained in this manner for a total of 7,576 Ap. mellifera genes.

Additional Analysis

Statistical tests (rank sum tests and correlations) were performed using either R (R Development Core Team 2010) or the JMP statistical software package (SAS Institute Inc.). Proportional Venn diagrams were generated using the Venn Diagram Plotter available from Pacific Northwest National Laboratory (http://omics.pnl.gov).

Results

We divided genes into low and high CpG_OE classes based on the bimodal distributions of CpG_OE observed in A. pisum (CpG_OE cutoff = 0.82; fig. 1A) and Ap. mellifera (CpG_OE cutoff = 0.72; fig. 1B). These two classes of genes roughly correspond to genes incurring relatively dense versus relatively sparse methylation (Saxonov et al. 2006; Suzuki et al. 2007; Weber et al. 2007; Elango et al. 2009; Foret et al. 2009; Wang and Leung 2009; Yi and Goodisman 2009; Xiang et al. 2010).

To gain insight into the evolutionary maintenance of genes with different levels of methylation, we first investigated whether genes belonging to distinct CpG_OE classes showed differences in their conservation of CpG_OE status over evolutionary time. A total of 2,339 three-way orthologs were identified with nonzero CpG_OE values in A. pisum, Ap. mellifera, and D. melanogaster. By comparing the CpG_OE classification of orthologs in A. pisum and Ap. mellifera from this data, we found that genes with high CpG_OE exhibited considerably less conservation of CpG_OE status than genes with low CpG_OE (fig. 2, table 1; Pearson’s Chi-squared test with Yates’ continuity correction P = 0.0075). Thus, patterns of dense DNA methylation have been more conserved over evolutionary time than patterns of sparse DNA methylation in A. pisum and Ap. mellifera.

We next determined whether the differential conservation of low CpG_OE and high CpG_OE status was associated with differential conservation of nucleotide and amino acid sequence. We found that genes from the low CpG_OE class in A. pisum and Ap. mellifera both harbored significantly greater proportions of genes with detectable three-way orthologs than genes from the high CpG_OE class (table 2; Pearson’s Chi-squared test with Yates’ continuity correction P < 1 × 10⁻¹⁵). We also found that DNA sequence conservation was significantly higher between A. pisum and Ap. mellifera orthologs from the low CpG_OE class than orthologs from the high CpG_OE class (Kruskal–Wallis rank sum test P = 0.0003; fig. 3A, supplementary table S1, Supplementary Material online). Both of these results suggested
that densely methylated genes, as a whole, were considerably more conserved at the sequence level than sparsely methylated genes. However, in contrast to the results obtained from analysis of ortholog loss and DNA sequence identity, amino acid substitution rates among genes with detectable three-way orthologs were slightly higher among low CpG\textsubscript{O/E} genes than high CpG\textsubscript{O/E} genes (Kruskal–Wallis rank sum test $P = 0.0012$; fig. 3B and supplementary fig. S1 and tables S1 and S2, Supplementary Material online). Furthermore, an alternate analysis, presented in our supplementary material, also found that densely methylated genes with detectable orthologs exhibited slightly higher rates of amino acid substitution than sparsely methylated genes.

To investigate whether genes with different levels of methylation were associated with specific functions, we next tested for enrichment of GO biological process terms in 4,404 \textit{A. pisum} genes with \textit{D. melanogaster} orthologs. We found that functions related to cellular metabolic processes were overrepresented among low CpG\textsubscript{O/E} genes (table 3). In contrast, functions associated with cellular signaling, behavior, and environmental stimulus were overrepresented among high CpG\textsubscript{O/E} genes (table 3).

We also found that six of the top ten enriched functional terms for \textit{A. pisum} low CpG\textsubscript{O/E} genes were among the top ten enriched functional terms in \textit{Ap. mellifera} low CpG\textsubscript{O/E} genes (table 3; Elango et al. 2009). In contrast, only two of the top ten high CpG\textsubscript{O/E} functional enrichment terms...
Low CpG significantly associated with high CpG each demonstrated that differential gene expression was by development, sex, and female wing polyphenism, which demonstrated that differential gene expression was by development, sex, and female wing polyphenism, which were in agreement between A. pisum and Apis mellifera (table 3; Elango et al. 2009). Thus, the function of low CpG genes appears to be relatively conserved over evolutionary history.

Finally, we investigated whether CpG O/E measures were associated with patterns of gene expression among distinct phenotypic groups in A. pisum using microarray data for 1,347 genes (Brisson et al. 2007). We analyzed the degree of differential gene expression between developmental stages (development; 4th instar vs. adult), between sexes (sex; male vs. asexual female), between environmentally sensitive asexual female wing phenotypes (female wing polyphenism; winged vs. unwinged), and between genetically determined male wing phenotypes (male wing polymorphism; winged vs. unwinged).

Our results suggested that genes with low levels of DNA methylation exhibited complex, condition-specific regulation of gene expression: differential gene expression, when combined for all pairwise comparisons of alternate phenotypes, displayed a significant positive correlation with CpG O/E in A. pisum (Pearson’s product-moment correlation $P < 0.001$; table 4, fig. 4A). This signal was primarily driven by development, sex, and female wing polyphenism, which each demonstrated that differential gene expression was significantly associated with high CpG O/E (table 4; fig. 4A). Differential gene expression between male wing morphs was not significantly associated with CpG O/E in A. pisum, although the trend was in the same direction as the other tests (table 4, fig. 4A).

We also reanalyzed data linking gene expression to methylation levels in Ap. mellifera to illustrate that differential gene expression between caste phenotypes (Elango et al. 2009) and gene expression breadth (Foret et al. 2009) were also each associated with CpG O/E (fig. 4B and C). Specifically, genes with differential expression between Ap. mellifera queens and workers, and those expressed in few Ap. mellifera tissues, preferentially exhibited high CpG O/E. Overall, our results reveal that genes with condition-specific regulation are associated with higher CpG O/E and lower levels of DNA methylation than ubiquitously expressed genes in both A. pisum and Ap. mellifera.

### Discussion

#### Gene Evolution and DNA Methylation

We have reported distinct levels of conservation of DNA methylation status for orthologs with heavy methylation (low CpG O/E) and sparse methylation (high CpG O/E) in the pea aphid, A. pisum, and the honeybee, Ap. mellifera (fig. 2, table 1). In particular, a greater proportion of orthologs maintain low CpG O/E status than high CpG O/E status over evolutionary time. Thus, genes that were presumably densely methylated in the ancestor of A. pisum and Ap. mellifera were more likely to remain methylated through evolutionary time, whereas genes with sparse methylation were less likely to maintain their low methylation status.

Furthermore, we found that heavily methylated genes had a greater number of detectable orthologs and exhibited greater DNA sequence conservation than genes with sparse methylation (table 2; fig. 3A). In line with these results, a prior study also found that genes with signatures of methylation were enriched among orthologs that could be identified between distantly related taxa (Suzuki et al. 2007). Thus, heavily methylated genes, overall, appear to be more conserved at the sequence level than sparsely methylated genes. This observation is particularly striking because DNA methylation increases the occurrence of mutations at CpG sites and might be expected to lead to rapid DNA sequence divergence (Elango et al. 2008). One possible explanation for the observed trend, however, is that orthologs with consistently low CpG O/E over evolutionary history have

---

**Table 1**

Contingency Table of CpG O/E Conservation between Acyrthosiphon pisum and Apis mellifera

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A. pisum</td>
<td>864</td>
<td>437</td>
<td>66.4</td>
</tr>
<tr>
<td>Low CpG O/E genes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. pisum</td>
<td>633</td>
<td>405</td>
<td>61.0</td>
</tr>
<tr>
<td>High CpG O/E genes</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

_Note._—Conservation differs significantly between low CpG O/E genes and high CpG O/E genes (Pearson’s Chi-squared test with Yates’ continuity correction $P = 0.0075$).

**Table 2**

Ortholog Detection among Low CpG O/E and High CpG O/E Genes

<table>
<thead>
<tr>
<th>Acyrthosiphon pisum</th>
<th>Apis mellifera</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low CpG O/E</td>
<td></td>
</tr>
<tr>
<td>Three-Way Orthology</td>
<td>1,301</td>
</tr>
<tr>
<td>No Three-Way Orthology</td>
<td>3,309</td>
</tr>
<tr>
<td>Proportion with Three-Way Orthology (%)</td>
<td>28.2</td>
</tr>
<tr>
<td>High CpG O/E</td>
<td></td>
</tr>
<tr>
<td>Three-Way Orthology</td>
<td>1,038</td>
</tr>
<tr>
<td>No Three-Way Orthology</td>
<td>4,818</td>
</tr>
<tr>
<td>Proportion with Three-Way Orthology (%)</td>
<td>17.7</td>
</tr>
</tbody>
</table>

_Note._—Ortholog detection differs significantly between low CpG O/E genes and high CpG O/E genes (Pearson’s Chi-squared test with Yates’ continuity correction $P < 1 \times 10^{-15}$ for both A. pisum and Ap. mellifera, each analyzed separately).
fewer total CpG dinucleotides than methylated genes with intermediate CpG\textsubscript{OEG}, and thus do not incur new mutations at a comparable rate (Suzuki et al. 2009). Another possibility is that genes targeted by DNA methylation may be under greater functional constraint, as a class, than unmethylated genes.

Surprisingly, in contrast to our results from analysis of DNA sequence identity, we found that densely methylated genes with detectable orthologs may be under less constraint at the amino acid level than their sparsely methylated counterparts (fig. 3B and supplementary fig. S1, Supplementary Material online). Apparently, A. pisum and Ap. mellifera high and low CpG\textsubscript{OEG} genes that do not retain detectable orthologs in D. melanogaster differ more from each other, in terms of evolutionary constraint at the protein level, than do high and low CpG\textsubscript{OEG} genes with detectable orthologs (table 2 and supplementary tables S1 and S2, Supplementary Material online; fig. 3 and supplementary fig. S1, Supplementary Material online). It remains unclear why this may be the case, but our results suggest that different classes of genes may behave differently with respect to the interaction between selective constraints or mutability and methylation status.

### Gene Expression and DNA Methylation

In the present study, we add to the emerging view that genes with ubiquitous expression in insects are preferentially targeted by DNA methylation (Elango et al. 2009; Foret et al. 2009; Xiang et al. 2010). Specifically, genes with similar expression levels among phenotypic groups exhibit evolutionary signatures of significantly higher levels of DNA methylation than genes with differential expression between different classes of genes.
phenotypes in both *A. pisum* and *Ap. mellifera* (fig. 4A and B; Elango et al. 2009). Genes with ubiquitous expression among tissues are also preferentially targeted by DNA methylation in both *Ap. mellifera* (fig. 4C; Foret et al. 2009) and the silkworm, *Bombyx mori*, even though *B. mori* possesses only a partial complement of DNA methylation enzymes (Xiang et al. 2010). By comparison, genes with tissue-specific expression in *Ap. mellifera* (fig. 4C; Foret et al. 2009) and *B. mori* (Xiang et al. 2010), with caste-specific expression in *Ap. mellifera* (fig. 4B; Elango et al. 2009), and with differential expression between developmental stages, sexes, and polyphenic wing morphs in *A. pisum*, all exhibit lower levels of DNA methylation than their ubiquitously expressed counterparts (fig. 4A). Thus, sparse levels of DNA methylation are associated with flexibility in gene expression, either between polyphenic forms or different tissues.

Our results reveal that complex gene regulation is associated with low levels of DNA methylation in disparate insects. This finding may appear to contrast with the idea that DNA methylation plays an important role in the epigenetic regulation of phenotypic plasticity (Jaenisch and Bird 2003; Kucharski et al. 2008; Maleszka 2008). Indeed, our observations suggest that the primary targets of DNA methylation are those genes least likely to be implicated as leading to phenotypic variation. However, we cannot rule out the cooption of DNA methylation

<table>
<thead>
<tr>
<th>Table 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlations between <em>Acyrthosiphon pisum</em> Differential Gene Expression and CpG&lt;sub&gt;0/E&lt;/sub&gt;</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Pearson Product-Moment Correlation with CpG&lt;sub&gt;0/E&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean $</td>
<td>\log_2$ expression ratio$</td>
</tr>
<tr>
<td>Mean $</td>
<td>\log_2$ expression ratio$</td>
</tr>
<tr>
<td>Mean $</td>
<td>\log_2$ expression ratio$</td>
</tr>
<tr>
<td>Mean $</td>
<td>\log_2$ expression ratio$</td>
</tr>
<tr>
<td>Mean $</td>
<td>\log_2$ expression ratio$</td>
</tr>
</tbody>
</table>

*P < 0.05, ***P < 0.001, ****P < 0.0001.

Fig. 4.—Ubiquitously expressed genes exhibit higher levels of DNA methylation than genes with condition-specific expression. (A) Genes with a high degree of differential expression between groups exhibit significantly higher CpG<sub>0/E</sub> than genes with ubiquitous expression in *Acyrthosiphon pisum*. This relationship also holds true for (B) differential expression between *Apis mellifera* queen and worker castes (adapted from Elango et al. 2009). (C) Similarly, genes with a high degree of tissue specificity exhibit significantly higher CpG<sub>0/E</sub> than genes with ubiquitous expression among tissues in *Ap. mellifera* (adapted from Foret et al. 2009). Significance values represent Wilcoxon signed-rank tests in panels A and B and a Kruskal-Wallis rank sum test in panel C. Means and 95% confidence intervals are plotted. Horizontal dashed lines represent the mean CpG<sub>0/E</sub> for all genes in a given panel. Vertical gray lines represent bin cutoffs for classification of genes according to mean $|\log_2$ expression ratio$|$.  

for complex regulatory roles operating on a smaller number of loci.

Steps toward a Unified View of Intragenic Methylation

Recently, a unified view of the functional role of intragenic (vs. intergenic or promoter) DNA methylation in vertebrates and invertebrates has begun to emerge. For example, methylation of gene bodies in many vertebrates and invertebrates is associated with moderate gene expression levels (Zemach et al. 2010). Our data, obtained from microarray analyses, do not directly address overall levels of gene expression but instead address expression breadth among tissues or alternate phenotypic classes. We find that genes with high CpG\text{avg} measures possess an enriched aptitude for conditional expression associated with distinct tissues or alternate phenotypes. In contrast, genes with dense methylation exhibit a greater propensity for static levels of expression.

A recent mammalian study revealed that intragenic methylation limits the generation of alternate gene transcripts by masking intragenic promoters (Maunakea et al. 2010). This mechanism may explain why broadly expressed genes are subject to the highest levels of methylation in invertebrates: broadly expressed genes may be preferentially targeted by DNA methylation due to enhanced negative effects associated with alternate promoters at such loci. Importantly, the proposed link between intragenic methylation and the regulation of alternate transcription (Maunakea et al. 2010) suggests that different levels of methylation in distinct tissues or developmental stages could have important phenotypic consequences.

Finally, we note that our results do not apply to insect taxa that have heavily diminished methylation systems (Urieli-Shoval et al. 1982; Field et al. 2004). Instead, we suggest that DNA methylation is one of many tools that can be co-opted for the purposes of gene regulation in organisms that have retained a complete enzymatic toolkit for mediating DNA methylation.

Supplementary Material

Supplementary figure S1 and tables S1–S2 are available at Genome Biology and Evolution online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

We thank Sylvain Foret and Christina Grozinger for sharing data from previous investigations and Jungsun Park from the Yi laboratory for sharing code used to analyze CpG depletion. We also thank three anonymous reviewers for comments which helped to improve this manuscript. This work was supported by the U.S. National Science Foundation (grant numbers MCB-0950896 to S.V.Y. and M.A.D.G., DEB-0640690 to M.A.D.G. and S.V.Y., and DEB-1011349 to B.G.H., S.V.Y., and M.A.D.G.) and National Institute of Environmental Health Sciences (grant number 4R00ES017367 to J.A.B.).

Literature Cited


Associate editor: George Weinstock