

Comparative Analyses of DNA Methylation and Sequence Evolution Using *Nasonia* Genomes

Jungsun Park,^{§1} Zuogang Peng,^{†§2} Jia Zeng,² Navin Elango,^{‡,2} Taesung Park,¹ Dave Wheeler,³ John H. Werren,³ and Soojin V. Yi^{*,2}

¹Bioinformatics and Biostatistics Laboratory, Interdisciplinary Program in Bioinformatics, Department of Statistics, Seoul National University, Seoul, Korea

²School of Biology, Georgia Institute of Technology

³Department of Biology, University of Rochester

[†]Present address: Southwest University School of Life Science, Chongqing, China.

[‡]Present address: Dow AgroSciences LLC, Indianapolis, Indiana.

[§]These authors contributed equally to this work.

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

Associate editor: John Parsch

Abstract

The functional and evolutionary significance of DNA methylation in insect genomes remains to be resolved. *Nasonia* is well situated for comparative analyses of DNA methylation and genome evolution, since the genomes of a moderately distant outgroup species as well as closely related sibling species are available. Using direct sequencing of bisulfite-converted DNA, we uncovered a substantial level of DNA methylation in 17 of 18 *Nasonia vitripennis* genes and a strong correlation between methylation level and CpG depletion. Notably, in the sex-determining locus *transformer*, the exon that is alternatively spliced between the sexes is heavily methylated in both males and females, whereas other exons are only sparsely methylated. Orthologous genes of the honeybee and *Nasonia* show highly similar relative levels of CpG depletion, despite ~190 My divergence. Densely and sparsely methylated genes in these species also exhibit similar functional enrichments. We found that the degree of CpG depletion is negatively correlated with substitution rates between closely related *Nasonia* species for synonymous, nonsynonymous, and intron sites. This suggests that mutation rates increase with decreasing levels of germ line methylation. Thus, DNA methylation is prevalent in the *Nasonia* genome, may participate in regulatory processes such as sex determination and alternative splicing, and is correlated with several aspects of genome and sequence evolution.

Key words: *Nasonia*, DNA methylation, genome evolution, sex determination, molecular evolution, alternative splicing.

Introduction

DNA methylation refers to the addition of the methyl group to the C5 position of the cytosine base. In animals, the main targets of DNA methylation are cytosines when followed by guanine in 3'. For this reason, DNA methylation in animals is also often referred to as "CpG methylation." Because methyl-cytosines allow a genome to increase its information content beyond the capacity of the canonical four base system, they are sometimes referred to as the "fifth" base (Lister and Ecker 2009). Indeed, methylated and unmethylated DNA often exhibit distinctive activity differences. For example, in mammals, DNA methylation is associated with repressed chromatin state and silenced gene expression (Bird and Wolffe 1999).

It has been known for some time that DNA methylation occurs in various invertebrates (Regev et al. 1998; Field et al. 2004), but its functional significance and consequences to invertebrate genome evolution have not been extensively studied. This is due in part to the fact that DNA methylation is lacking or nearly absent in the two traditional invertebrate genetic model organisms, *Drosophila melanogaster* and *Caenorhabditis elegans*. However, with the technical advances in genome

sequencing and efficient detection of methyl-cytosines, functional DNA methylation has been recently confirmed in genome-wide surveys of several insects, including the honeybee (Wang et al. 2006), the pea aphid (Walsh et al. 2010), the silkworm (Xiang et al. 2010), the jewel wasp (Werren et al. 2010), and several ant species (Bonasio et al. 2010; Smith et al. 2011; Wurm et al. 2011). The finding that diverse invertebrates have DNA methylation indicates that it is ancestral and that loss of extensive DNA methylation is derived in some taxa, such as flies. Additionally, studies indicate that the most conserved pattern of DNA methylation in invertebrates is that of transcribed genic regions or "gene bodies" (Feng et al. 2010; Zemach et al. 2010). In contrast, methylation of promoters is vertebrate specific (Elango and Yi 2008).

These discoveries have invigorated interest in understanding the function and evolution of DNA methylation in invertebrate taxa. Hymenopteran insects are emerging as useful model systems to investigate these topics. For example, recent studies using the honeybee *Apis mellifera* have begun to shed light on the importance of DNA methylation in hymenopteran species. The honeybee genome

encodes a complete suite of methylation enzymes and harbors a functional DNA methylation system (Wang et al. 2006). Similar to other invertebrates, DNA methylation in the honeybee is targeted toward genes rather than nongenic regions (Wang et al. 2006). The level of methylation varies greatly among the honeybee genes. It has been demonstrated, computationally and experimentally, that honeybee genes are divided into two distinctive groups according to the levels of DNA methylation: densely and sparsely methylated genes (Elango et al. 2009; Foret et al. 2009; Wang and Leung 2009; Lyko et al. 2010; Zemach et al. 2010). Genes sparsely methylated in the germ lines tend to be preferentially involved in caste-specific patterns of gene expression (Elango et al. 2009). Densely and sparsely methylated genes are also distinct in levels of gene expression (Foret et al. 2009) and exhibit contrasting enrichments in gene function categories (Elango et al. 2009; Wang and Leung 2009). Gene lengths, in particular introns lengths, also appear to differ greatly between sparsely and densely methylated honeybee genes (Zeng and Yi 2010). These studies provide interesting directions for future investigations, including comparative studies of methylation.

The newly sequenced genomes of the jewel wasp *Nasonia vitripennis* and its sibling species (Werren et al. 2010) provide a unique opportunity to investigate evolutionary patterns of genomic DNA methylation and its influence on genome evolution. *Nasonia* are non-eusocial wasps whose larvae parasitize other arthropods. Due to their ability to specifically target host species, parasitoid wasps have a great potential to be used as biological control agents against insect pests. Importantly, unlike the honeybee, *Nasonia* is highly amenable to the common laboratory environments, making it a tractable model to investigate functional roles of DNA methylation. Like *A. mellifera*, the genome of *N. vitripennis* encodes a complete protein suite required for functional DNA methylation (Werren et al. 2010). *Nasonia* are well positioned for comparative analyses of DNA methylation due to the availability of genomic resources from a moderately divergent outgroup (the honeybee, which has diverged from *Nasonia* ca. 190 Ma; Werren et al. 2010) as well as closely related sibling species in the genus *Nasonia* (Werren et al. 2010).

Here, we present analyses of DNA methylation in *Nasonia* genomes. We uncovered substantial levels of DNA methylation from selected genes and intriguing variation of relative CpG depletion levels. Furthermore, comparative analyses using distant and closely related outgroups allow us to delineate evolutionary patterns of DNA methylation in relation to genome evolution in different evolutionary timescales.

Materials and Methods

Experimental Detection of Methyl-Cytosines

To quantify the level of DNA methylation of selected *N. vitripennis* genes, we used direct sequencing of polymerase chain reaction (PCR) products following bisulfite conversion of genomic DNA. Total genomic DNA was isolated using Pure-gene DNA isolation kit (Gentra/Qiagen) from 10 to 20 pooled individuals each from the following six experimental groups: males and females for three distinctive developmental stages

(adult, yellow, and black pupae). Genomic DNAs were then aliquoted to 500 ng each and then bisulfite converted using the EpiTect Bisulfite conversion kit (Qiagen) following the manufacturer's instructions.

We designed bisulfite sequencing primers using the Methyl Primer Express Software (v1.0) (Applied Biosystems). We initially developed 72 primer pairs based upon 41 genes selected from the *Nasonia* RefSeq gene set (shown in [supplementary table 1, Supplementary Material](#) online). These 41 genes were selected to represent similar numbers of low- and high-CpG O/E groups of genes. However, because it is harder to design primers for high-CpG O/E group genes (because it is difficult to choose priming sites with no CpG dinucleotides), we ended up with 25 low-CpG O/E genes and 16 high-CpG O/E genes.

Each primer pair was then amplified in a 25 μ l reaction PCR reaction. In our initial PCR screen, 44 primer sets (of 72 primer pairs tested) produced strong reproducible bands of correct sizes. We then amplified these from the six experimental groups (two sexes for three developmental stages). Some of the primers did not amplify from all six groups and were not included in the subsequent steps. For some genes, we designed several pairs of primers to increase the chance to succeed in PCR. When more than one primer pairs produced PCR products from the same genomic regions, we chose the ones resulted in the brightest PCR bands.

Amplified bisulfite PCR products were purified using the QIAquick Gel Extraction Kit (Qiagen), then cloned into pCR 2.1 vector by use of a TOPO-TA cloning system (Invitrogen), transformed into TOPO10 chemically competent *Escherichia coli* (Invitrogen). We diluted samples 1,000 times before plating to increase the probability of unique alleles. This approach provides an approximation on the level of DNA methylation (Farcas et al. 2009). A least five (most of them greater than eight) positive clones were randomly selected for Sanger sequencing. Following these steps, we collected data from 25 primer pairs, distributed over 18 genes (table 1).

Measurement and Classification of CpG O/E Distribution

CpG O/E or "normalized CpG content" is a metric of depletion of CpG dinucleotides. It is negatively correlated with DNA methylation levels in diverse animal genomes (Bird and Taggart 1980; Elango and Yi 2008; Xiang et al. 2010; Zemach et al. 2010).

CpG O/E is defined as

$$\begin{aligned} \text{CpG}[O/E] &= \frac{P_{\text{CpG}}}{P_C \times P_G} \\ &= \frac{\text{number}(\text{CpG})}{\text{number}(\text{C}) \times \text{number}(\text{G})} \\ &\quad \times \frac{(\text{length})^2}{\text{length}}, \end{aligned}$$

where P_{CpG} , P_C , and P_G are the frequencies of CpG dinucleotides, C nucleotides, and G nucleotides, respectively. We calculated CpG O/E for each gene, using RefSeq annotations. We analyzed intron data as well as data from gene bodies (exons + introns).

Statistical Analyses of CpG O/E Distribution in *N. vitripennis* Genes

In the jewel wasp, the distributions of CpG O/E from gene bodies (defined as exons + introns, as in; Suzuki and Bird 2008; Zemach et al. 2010) as well as introns can be described as a mixture of several distributions (fig. 2, Supplementary text, Supplementary Material online). We estimated the number of components in those mixture distributions using a model-based clustering. The mclust package (Fraley and Raftery 2003) in R package (www.r-project.org) was used to estimate the number of components under the Gaussian Mixture Model. This model is described as

$$\sum_{i=1}^k p_i N(\mu_i, \sigma_i),$$

where the function N is a Gaussian model containing unknown parameters μ_i (mean of each component) and σ_i (standard deviation of each component). k is the number of components in the mixture model. p_i is the proportion of each Gaussian model component in the mixture model. These parameters are estimated using the Expectation–Maximization algorithm.

Previously, we used a likelihood-ratio test (LRT) to compare the fit of a unimodal ($k = 1$) to that of a bimodal model ($k = 2$) (Elango et al. 2008, 2009; Hunt et al. 2010). LRT is limited since it allows comparison of nested models only. Here, to identify the best fitting model among several non-nested models, we used Akaike information criteria (AIC) and Bayesian information criteria (BIC) (Supplementary text, Supplementary Material online). Using the mclust package also allows us to apply more stringent criteria for assigning different components (Supplementary text, Supplementary Material online).

Ortholog Identification and Estimation of Nucleotide Substitution Rates

Orthologs between *N. vitripennis*, *A. mellifera*, and *D. melanogaster* were identified in Werren et al. (2010). Briefly, all-against-all protein sequence comparisons were performed using the Smith–Waterman algorithm. Best reciprocal hits were then clustered using a cutoff E value of 10^{-6} . The numbers of nonsynonymous and synonymous substitutions between the honeybee and the jewel wasp as well as between the three sibling species of *Nasonia* were estimated using the codeml module of the PAML package (Yang 2007).

To estimate divergence data from introns, we first generated multiple sequence alignments of orthologous intronic regions from *N. vitripennis*, *N. giraulti*, and *N. longicornis* by ClustalW using the default settings. A computational pipeline was used to identify and remove low-quality alignments or regions that contained mostly missing data (N 's) before further analyses. Maximum likelihood estimates of divergence (dI) were obtained from the intron alignments using the baseml module of the PAML package (Yang 2007). The model of sequence evolution used for the baseml calculations was HKY85 and no molecular clock was assumed (clock = 0).

Gene Ontology Enrichment Analyses

We investigated overrepresentation of specific gene ontology (GO) terms in genes belonging to different CpG O/E groups. Due to the lack of extensive functional annotations of *N. vitripennis* genes, we used 1:1 orthologs to *D. melanogaster* and used GO terms in *D. melanogaster* orthologs for this analysis. We used the GeneTrail tools (<http://genetrail.bioinf.uni-sb.de/>) to test for enrichments of specific GO biological process terms, using the whole orthologs as the background set. Statistical significances are calculated using a Hypergeometric distribution and correcting for multiple testing by false discovery rates adjustment.

Results

DNA Methylation Is Prevalent in *N. vitripennis* Genome

To examine the presence and patterns of DNA methylation in *N. vitripennis* genome, we experimentally characterized CpG methylation using sequencing of bisulfite-converted genomic DNA. Briefly, sodium bisulfite converts unmethylated cytosine residues to uracil (which is subsequently converted to thymine during PCR), whereas methylated cytosines remain unmodified. Thus, methylated and unmethylated cytosines can be distinguished by comparing sequences from bisulfite-treated and not-treated DNA. In other words, we can map DNA methylation in a nucleotide-specific manner.

We analyzed DNA methylation in genes because DNA methylation is known to preferentially target gene bodies, especially exons, in insects as well as other invertebrates (Wang et al. 2006; Suzuki et al. 2007; Xiang et al. 2010; Zemach et al. 2010). Table 1 lists 18 genes we surveyed for the evidence of DNA methylation. Among these 18 selected genes, all but one showed the presence of methyl-cytosines. We did not detect any non-CpG methylation from this analysis, suggesting that the level of non-CpG methylation may be negligible compared with CpG methylation. This is similar to what has been observed in the honeybee (Wang et al. 2006).

Even though we examined six different experimental groups (two sexes and three developmental stages), we did not observe any sex- and developmental stage-specific patterns of DNA methylation in these genes. Consequently, we present the means of all six groups in table 1. We estimated the number of methylated cytosines by counting the numbers of CpGs harboring methylated cytosines relative to all CpGs in the examined samples. These ratios, ${}^m\text{CG}/({}^m\text{CG} + \text{CG})$, are shown in table 1. The mean ${}^m\text{CG}/({}^m\text{CG} + \text{CG})$ ratio of the 18 genes is 0.315. Thus, a considerable number of methyl-cytosines is present in *N. vitripennis* genome. However, we caution that our sampling scheme is not random, and this value cannot necessarily be extrapolated into the genome-wide level. Our initial primer sets included more from the low-CpG genes, which are more likely to be methylated. In addition, only those genes that we were able to amplify from all samples using PCR were included in the survey. Nonbiased sampling scheme such as sequencing of bisulfite-converted genomic DNA should be used to determine the genome-wide levels of DNA methylation in *N. vitripennis*.

Table 1. Eighteen Genes Sequenced by Using Bisulfite Sequencing in the Genome of *Nasonia vitripennis*.

Gene ID		CpG O/E	Number of CpG sites	Amplicon Length	^m CG/(^m CG + CG)	
XM_001600728.1	Epithelial membrane protein	exon 3	1.579	29	357	0.027
XM_001607338.1	Vitellogenin	exon 3	1.002	23	360	0.006
XM_001601674.1	Odorant receptor 85c	exon 3	1.218	14	297	0.009
XM_001605749.1	Integrator complex subunit 6	exon 5	1.216	24	303	0.003
XM_001604744.1	Transformer		1.171	47	1171	0.092
	Amplicon 1	exon 2		7	301	0.701
	Amplicon 2	exon 3–4		9	289	0.012
	Amplicon 3	exon 5–6		13	255	0.007
	Amplicon 4	exon 7–8		18	326	0.011
XM_001602423.1	Fruitless gene	exon 3	1.144	18	309	0.004
XM_001599168.1	Proteasome beta 3 subunit	exon 2	0.343	3	284	0.882
XM_001600593.1	Polypeptide of 976 aa		0.459	16	992	0.739
	Amplicon 1	exon 2		7	337	0.948
	Amplicon 2	exon 3		6	342	0.461
	Amplicon 3	exon 4		3	313	0.642
XM_001606530.1	eIF 2a kinase		0.420	7	703	0.571
	Amplicon 1	exon 2–3		2	364	0.534
	Amplicon 2	exon 4		5	339	0.645
XM_001601041.1	eIF2B-gamma protein		0.362	7	657	0.829
	Amplicon 1	exon 2		2	393	0.829
	Amplicon 2	exon 3		5	264	0.827
XM_001599009.1	Enhancer of zeste, ezh	exon 2	0.796	5	385	0.852
XM_001600017.1	ATP-dependent RNA helicase	exon 3	0.797	5	334	0.379
XM_001601077.1	Retrotransposon protein	exon 4	0.792	6	245	0.003
XM_001601755.1	Cytochrome P450	exon 3	0.778	10	345	0.007
XM_001602655.1	Ras gtpase activating protein	exon4	0.783	8	325	0.500
XM_001603232.1	Retrotransposon protein	exon 3	0.799	4	317	0
XM_001605357.1	Glutamate cysteine ligase	exon 6	0.791	10	287	0.05
XM_001605741.1	Slender lobes	exon 3	0.790	7	326	0.717

In one of the genes, the *transformer* locus, data included three intronic regions (fig. 1). Amplicons in other loci did not include any intronic CpG dinucleotides. The ^mCG/(^mCG + CG) for the intronic sites is 0.008, much lower than other sites. This observation is in accord with the idea that exons are more strongly methylated than introns, similar to the case in the honeybee (Wang et al. 2006).

The data in table 1 provide a chance to examine the relationship between CpG O/E and the level of DNA methylation in *N. vitripennis*. CpG O/E, a metric of CpG depletion (Materials and Methods), is used as a proxy for the levels of DNA methylation in diverse animal taxa (e.g., Suzuki et al. 2007; Weber et al. 2007; Elango and Yi 2008). Briefly, DNA methylation occurs predominantly at CpG dinucleotides in animal genomes. Methylated cytosines are chemically unstable and change to thymine via spontaneous deamination. This leads to increased frequency of CpG to TpG (CpA in the complementary strand) mutations. Consequently, methylated regions gradually lose CpG dinucleotides (Bird 1980; Duncan and Miller 1980), and CpG O/E is expected to be negatively correlated with the levels of DNA methylation. Indeed, recent surveys of genome-wide methylation, including in the honeybee (Lyko et al. 2010) and the silkworm (Xiang et al. 2010), provide supports for the use of CpG O/E as an indicator of the methylation level in insect genomes.

In *N. vitripennis*, CpG O/E and ^mCG/(^mCG + CG) ratios are highly significantly negatively correlated among the 18 genes that we examined (Pearson's $r = -0.72, P < 10^{-3}$). Thus, CpG O/E measures appear to be a good indicator of the level of DNA

methylation in this genome. We also observe a similar correlation between CpG O/E and experimentally measured methylation from the honeybee (S. Sarda, J. Zeng, B. Hunt, S. Yi, unpublished data). It is important to note that because CpG O/E measures depletion of CpG dinucleotides at inherited DNA, it best reflects levels of germ line methylation. Nevertheless, studies in mammals show that methylation levels in different tissues tend to be strongly correlated, and CpG O/E is consequently highly related to levels of DNA methylation in whole bodies and in other, non-germ line tissues (Weber et al. 2007; Bock and Lengauer 2008).

Differential Methylation Among the Exons of the *transformer* Locus

To determine whether there is intragenic variation in the levels of DNA methylation, we further analyzed patterns of DNA methylation of multiple exons for 4 of the 18 genes (RefSeq IDs XM_001604744.1, XM_001600593.1, XM_001606530.1, and XM_001601041.1, table 1). Among the four genes analyzed, the major sex-determining locus *transformer* (*tra*) shows exon-specific differences in methylation level (fig. 1). This locus (referred to a *feminizer* in *Apis*, Hasselmann et al. 2008) is homologous to the splicing factor *transformer*, which has a conserved role as a sex-determining signal by regulating female specific-splicing of “doublesex” (Sosnowski et al. 1989). Although exons 3–8 are only weakly methylated (^mCG/(^mCG + CG) ratios are around 1%), the second exon exhibits high frequency of DNA methylation (^mCG/(^mCG + CG) was 0.7 for the second exon). The results are

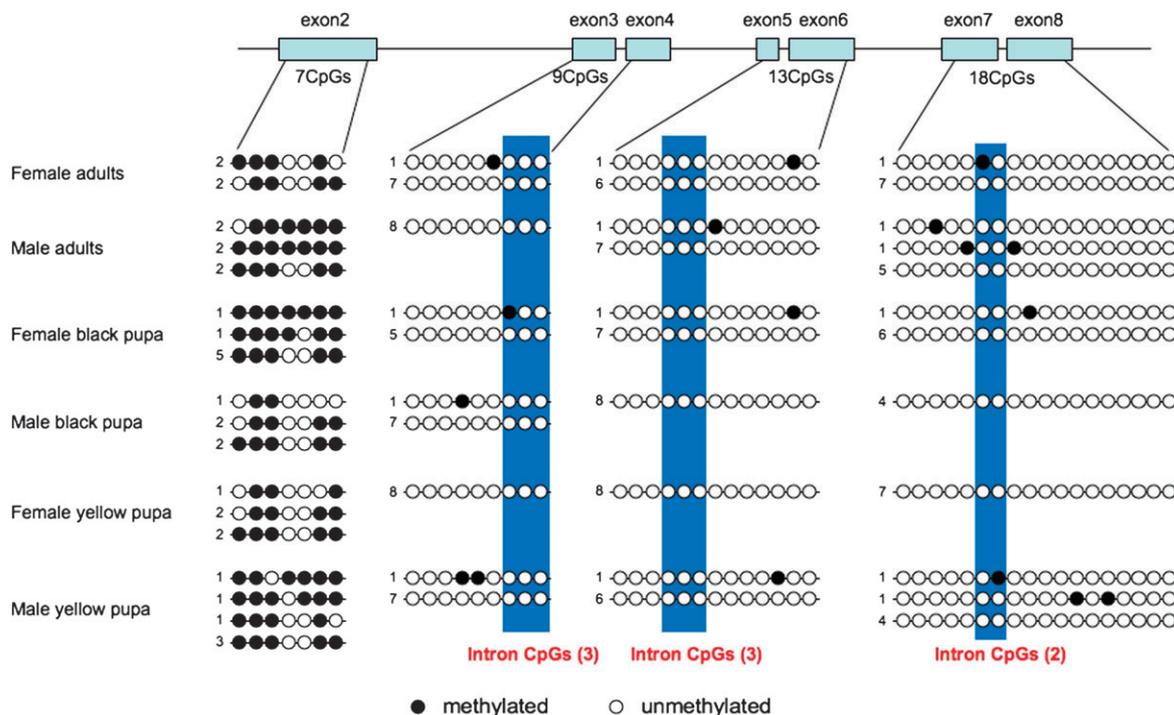


Fig. 1. Complex pattern of DNA methylation uncovered by experimental means from multiple exons of the *transformer* locus. CpG O/E calculated from the whole gene is 1.171, suggesting a generally sparse level of DNA methylation. Accordingly, CpG dinucleotides in exons 3–8 show sparse levels of DNA methylation in all developmental stages and sex surveyed. However, CpG dinucleotides in the exon 2 are densely methylated in most cases. There are eight CpG dinucleotides belonging to introns (highlighted in blue). The frequency of intronic methylcytosines is much lower than that of exons.

intriguing as exon 2 shows differential splicing between male- and female-specific transcripts (Verhulst et al. 2010). Note that our data did not reveal a difference in methylation of this exon between males and females. The other three loci showed similar levels of DNA methylation among different exons.

Functional Enrichments and Gene Lengths Differ Between Low- and High-CpG Genes

To infer the genomic pattern of DNA methylation, we analyzed distributions of CpG O/E in *N. vitripennis* genes. We analyzed introns as well as gene bodies (exons + introns), as gene bodies are primary targets of DNA methylation in invertebrate genomes (Feng et al. 2010; Xiang et al. 2010; Zemach et al. 2010). We used the AIC and BIC to test the fit of mixture distributions with different numbers of components (see Materials and Methods and supplementary material, [Supplementary Material](#) online).

In the case of the honeybee or pea aphid, genes were clearly separated into two groups according to levels of CpG depletion (Elango et al. 2009; Walsh et al. 2010). This was not the case in *Nasonia*, where the levels of CpG depletion exhibited a relatively broad distribution (fig. 2). The pattern of CpG depletion in *Nasonia* is more similar to that observed in the pacific oyster *Crassostrea gigas* (Gavery and Roberts 2010). Likewise, fitting mixture distributions indicate that a model with two components can describe pattern of CpG depletion in *N. vitripennis*, for both intron and gene body data (fig. 2, [supplementary text](#), [Supplementary Material](#) online). Over

80% of genes are classified to the same components regardless whether introns or gene bodies are used. Following the terminologies in literature (Elango et al. 2009; Walsh et al. 2010), we refer to the two groups as low-CpG O/E genes and high-CpG O/E genes.

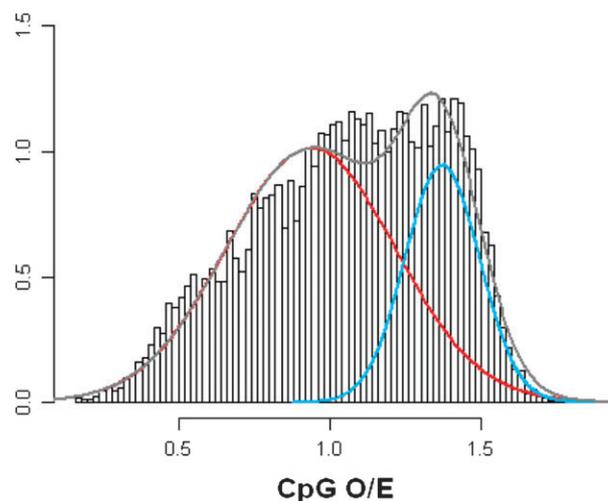


Fig. 2. Density plots of CpG O/E values from gene bodies of the jewel wasp, *Nasonia vitripennis*. Genes in the jewel wasp can be divided into two distinctive groups, referred to as low- and high-CpG O/E genes (red and blue curves, respectively). Gray lines represent the sum of the inferred components or the fit of the model to the observed distribution.

Table 2. Distinctive Functional Enrichments of the Low- and High-CpG O/E Genes.

GO Biological Process Term	Expected	Observed	Significance
Low-CpG O/E genes			
Cellular macromolecule			
metabolic process	575.1	678	1.0×10^{-10}
RNA processing	98.7	146	1.2×10^{-9}
Translation	97.3	138	2.7×10^{-7}
Gene expression	375.7	448	4.0×10^{-7}
Cellular protein metabolic			
process	285.4	349	7.1×10^{-7}
mRNA metabolic process	81.3	115	4.5×10^{-6}
Macromolecule metabolic			
process	659.3	737	5.6×10^{-6}
mRNA processing	75.6	107	1.1×10^{-5}
Cellular metabolic process	786.4	864	1.2×10^{-5}
High-CpG O/E genes			
System development	191.6	304	5.2×10^{-22}
Multicellular organismal			
development	237.8	350	7.2×10^{-19}
Organ development	147.7	241	1.1×10^{-18}
Transcription factor activity	49.5	105	2.9×10^{-17}
Anatomical structure			
development	235.7	341	7.9×10^{-17}
Developmental process	267.4	376	1.7×10^{-16}
Signal transducer activity	57.7	115	3.7×10^{-16}
Signaling	173.5	261	8.3×10^{-15}
Sequence-specific DNA			
binding	30.5	71	1.0×10^{-14}

We then investigated whether the two CpG O/E groups of genes in the jewel wasp exhibit distinctive functional enrichment, by analyzing overrepresentation of GO terms of the two groups. For subsequent analyses, we only used genes that belong to the same CpG O/E groups in both intron and gene body classifications. In the honeybee, low- and high-CpG genes exhibit strikingly different functional enrichments (Elango et al. 2009). Specifically, low-CpG genes are overrepresented in “housekeeping” processes, such as metabolic process and nucleotide processing. High-CpG groups, in contrast, are enriched in “developmental” functions. Remarkably, low- and high-CpG genes of the jewel wasp exhibit distinctive functional enrichment similar to the pattern observed in the honeybee (table 2). Specifically, low-CpG genes are overrepresented by functions in basic cellular processes, such as nucleotide processing, transcription, and biogenesis. High-CpG groups of genes are heavily enriched in development and morphogenesis functions. Such pattern of enrichment was the same when we used gene body or intron classifications separately (results not shown).

We have previously shown that low- and high-CpG genes in honeybee tend to be short and long, respectively (Zeng and Yi 2010). We observe the same pattern in *N. vitripennis*: CpG O/E is highly correlated with the length of gene bodies (Spearman's $r = 0.59, P < 10^{-16}$) as well as the length of introns (Spearman's $r = 0.49, P < 10^{-16}$). Exon lengths are also highly significantly correlated with CpG O/E (Spearman's $r = 0.13, P < 10^{-16}$), although not as strongly as intron or gene body lengths are. The mean intron lengths of low-CpG and high-CpG genes are also highly significantly different from each other (3,420 and 16,681 nt for low- and high-CpG genes, respectively,

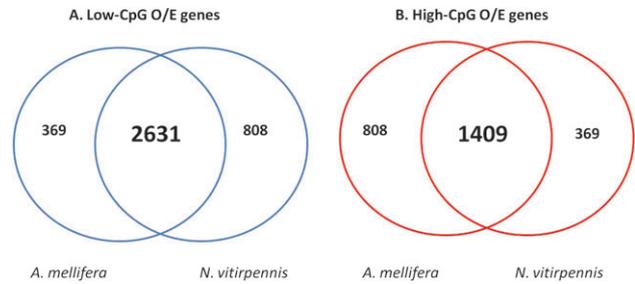


Fig. 3. Distributions of low- and high-CpG genes among honeybee and jewel wasp orthologs. Each circle represents the number of genes belonging to the corresponding CpG O/E status in either *Apis mellifera* or *Nasonia vitripennis*. The overlaps designate the numbers of orthologs from both species with the same CpG O/E classification. The low-CpG genes are depicted in (A), and the high-CpG genes are shown in (B). The low-CpG O/E genes are more conserved than the high-CpG O/E genes.

$P < 10^{-16}$). Total gene lengths and exon lengths also show a similar pattern (the means of low- and high-CpG O/E genes are 4,606 vs. 18,553 nt and 1,921 and 2,166 nt for gene bodies and exons, respectively, $P < 10^{-16}$ for both comparisons). Thus, the covariation between DNA methylation, functional enrichment, and gene lengths difference is a common theme in *A. mellifera* and *N. vitripennis*.

Evolutionary Conservation of DNA Methylation Status

To gain insights into the evolutionary maintenance of genes with respect to the levels of DNA methylation, we investigated evolutionary conservation and divergence of CpG O/E groups between the honeybee and the jewel wasp. *Nasonia* and the honeybee are estimated to have diverged 180–200 Ma (Werren et al. 2010). The CpG O/E values of orthologous genes between the jewel wasp and the honeybee are highly significantly correlated with each other (Spearman's $r = 0.64, P < 10^{-16}$). This demonstrates that DNA methylation levels of orthologous genes tend to be conserved between these two species.

We then examined the correspondence between different CpG O/E groups of the honeybee and the jewel wasp in detail. Most (77.4%) of the 5,217 1:1 ortholog genes between *N. vitripennis* and *A. mellifera* are classified into the same CpG O/E classes in the two species (fig. 3).

However, the proportions of genes belonging to the same CpG O/E class differed between low- and high-CpG O/E groups of genes. Among the low-CpG O/E group of 1:1 orthologous genes, 88% of genes remained in the same CpG O/E class in the two species. In contrast, 64% of genes stayed in the high-CpG O/E class in the two species. This difference is highly significant (Fisher's exact test, $P < 10^{-16}$, table 3). Thus, low-CpG O/E status tends to be more conserved compared with the high-CpG O/E status. Furthermore, the correlation of CpG O/E of orthologous genes between the honeybee and *N. vitripennis* is much stronger in low-CpG O/E group (Spearman's $r = 0.42, P < 10^{-16}$) than in high-CpG O/E group (Spearman's $r = 0.22, P < 10^{-16}$). In conclusion, densely methylated genes tend to stay densely methylated during the divergence time between the honeybee and the jewel wasp.

Table 3. Differential Conservation of CpG O/E Status Between Low- and High-CpG Genes in *Nasonia vitripennis*.

	Conserved CpG Group with <i>Apis mellifera</i>	Nonconserved CpG Group with <i>A. mellifera</i>	Proportion Conserved (%)
<i>N. vitripennis</i> genes			
Low-CpG O/E genes	2,631	369	88
High-CpG O/E genes	1,409	808	64
	P value (Fisher's exact test)		<10 ⁻¹⁶

NOTE.—CpG O/E values of *N. vitripennis* divide into low- and high-CpG genes, representing densely and sparsely methylated genes, respectively. Genes belonging to the low-CpG O/E class in *N. vitripennis* exhibit significantly greater conservation of CpG O/E status compared with high-CpG O/E genes.

Signature of DNA methylation Covaries with Rates of Nucleotide Substitution

Genome sequences for two sibling species of *N. vitripennis* (*N. longicornis* and *N. giraulti*) allowed us to examine molecular evolutionary patterns of genes with respect to CpG O/E status, by comparing substitution patterns in coding genes relative to *N. vitripennis* as the outgroup (Werren et al. 2010). These species are estimated to have diverged from *N. vitripennis* approximately 1 Ma and from each other approximately 0.5 Ma (Raychoudhury et al. 2009). *Nasonia longicornis* and *N. giraulti* show an average synonymous divergence of 0.03 to *N. vitripennis* and 0.012 to each other (Raychoudhury et al. 2009; Werren et al. 2010). We first examined the relationship between the rate of nonsynonymous, synonymous, and intronic divergence to CpG O/E.

According to the mutational tendency of methylated cytosines (Duncan and Miller 1980), genes that are subject to more methylation should exhibit an increased tendency toward mutation. Thus, nucleotide substitution rates might decrease with CpG O/E. In contrast with this expectation, we found that nonsynonymous, synonymous, and intronic rates are all positively correlated with CpG O/E (table 4). In other words, genes that are likely to be less methylated appear to have higher effective mutation rates. The fact that intronic rates exhibited the same pattern suggests that the effect is not due to functional constraints in the coding regions. Selective constraint (dN/dS) is not significantly correlated with CpG O/E (table 4).

We also examined the relationship between nucleotide substitution rates and CpG O/E on longer timescales by comparing orthologs of honeybee and *Nasonia*. Because of the long evolutionary timescales (~190 My divergence), most genes are saturated for synonymous substitutions. We therefore only present correlations between rates of nonsynonymous substitution (dN) and CpG O/E. Again, we observe a significant positive relationship between CpG O/E and rates of nucleotide substitution, although it is very weak (table 4).

We investigated whether the observed patterns can be caused by confounding effects of nucleotide contents and/or functional constraints. CpG O/E and G+C contents are correlated in many genomes (e.g., Elango et al. 2008; Zeng and Yi 2010), and G+C contents are often positively correlated with evolutionary rates (Yi et al. 2002). Thus, we

Table 4. Correlations Between Evolutionary Rates and CpG O/E.

	CpG O/E	P Value	CpG O/E GC Content	P Value
Comparison among <i>Nasonia</i> sibling species				
dN	0.117	<10 ⁻¹⁶	0.153	<10 ⁻¹⁶
dS	0.218	<10 ⁻¹⁶	0.157	<10 ⁻¹⁶
dI	0.226	<10 ⁻¹⁶	0.130	<10 ⁻¹⁶
dN/dS	-0.002	NS	-0.001	NS
<i>Nasonia</i> – <i>Apis</i> comparison ^a				
dN	0.066	<10 ⁻⁵	0.099	<10 ⁻¹¹

NOTE.—Comparisons in two timescales (between sibling species of *Nasonia* and between *Nasonia* and honeybee) provide support for a negative relationship between the level of DNA methylation and rates of nucleotide substitution.

^a Because synonymous sites are saturated for most genes, we only analyze nonsynonymous substitutions.

asked if the relationship between CpG O/E and substitution rates could be explained by the underlying G+C contents. We calculated partial correlations between CpG O/E and substitution rates, after removing the contributions of G+C contents (Kim and Yi 2007). The results are still highly significant, indicating that covariation between CpG O/E and G+C contents cannot account for the observed relationships between CpG O/E and substitution rates in all comparisons (table 4). In fact, the partial correlations between CpG O/E and dN increase after removing the effects of G+C contents.

Alternatively, because densely methylated (low-CpG O/E) genes tend to be enriched in housekeeping functions (table 2), it is possible that the correlation between CpG O/E and sequence divergence reflects bias toward more conserved genes in the low-CpG O/E category. To control for a potential effect of functional bias, we performed the following analysis. We first identified GO terms of orthologous proteins between *N. vitripennis* and *D. melanogaster*. We then examined the correlations between CpG O/E and sequence divergence among genes within each GO term. We limited our analyses to GO terms with at least 200 orthologous proteins, to increase statistical power. If the significant correlations between sequence divergence and CpG O/E are due to the functional bias, such correlation would disappear within each GO term. In contrast, we observe that in many cases, sequence divergence and CpG O/E are highly significantly correlated (table 5). In particular, the correlation between intron divergence and CpG O/E is highly significant in most (nine of ten) GO terms examined. Nonsynonymous rates (dN) and synonymous rates (dS) are significantly correlated with CpG O/E in five and six GO terms of ten terms, respectively. The fact that the correlation between rates of intron divergence and CpG O/E is a consistent pattern in this analysis again supports the idea that mutation rate differences underlie the relationship between CpG O/E and sequence divergence.

Discussion

The genome of *N. vitripennis* harbors a full set of methyltransferases required for maintenance and de novo DNA methylation (Werren et al. 2010). Our survey of 18 genes confirms that cytosine methylation is present in this genome and that the frequency of methyl-cytosines is likely to be

Table 5. Correlations Between CpG O/E and Sequence Divergence within Each GO Term.

GO Terms	Correlation with CpG O/E		
	dN	dS	dI
Regulation of transcription	0.20**	0.26***	0.16**
Cellular component morphogenesis	0.12 ^{NS}	0.17**	0.21**
Phosphate metabolic process	0.06 ^{NS}	0.24**	0.17**
Cytoskeleton organization	-0.04 ^{NS}	0.06 ^{NS}	0.21**
Cell morphogenesis	0.15**	0.16**	0.19**
Transcription	0.14**	0.24**	0.04 ^{TS}
Protein localization	0.15**	0.02 ^{NS}	0.14**
Neuron differentiation	0.27**	0.31***	0.21**
Microtubule-based process	0.03 ^{NS}	0.11 ^{NS}	0.22**
Postembryonic development	0.08 ^{NS}	0.11 ^{NS}	0.18**

NOTE.—For this analysis, we used data from *Nasonia* sibling species.
^{NS} $P > 0.05$; ** $P < 10^{-3}$; and *** $P < 10^{-5}$.

substantial: Among all CpG dinucleotides examined, 31.5% are in the form of methyl-cytosines (table 1). To compare the level of methylation in the jewel wasp genome to that of the honeybee genome, we performed the following analysis. Among the 18 genes we investigated here, ten genes had orthologs in the honeybee genome associated with experimentally determined methylation levels (Zemach et al. 2010). Among these ten genes, there was no difference in terms of actual methylation levels between the two species (Wilcoxon rank-sum test, $P = 0.35$). Thus, there is no evidence to suggest that *N. vitripennis* and *A. mellifera* have significantly different levels of genomic DNA methylation. It is also notable that the mean levels of CpG depletion in the *Nasonia* and the honeybee genome are highly similar (mean CpG O/E = 1.02 and 1.05 for the honeybee and *N. vitripennis*, respectively).

Remarkably, in one gene (*transformer*), methylation preferentially targets CpG sites of a particular exon. This pattern is suggestive of a role for DNA methylation in sex determination in *Nasonia* for two reasons. First, exon 2 shows differential splicing between male- and female-specific transcripts, and second, maternal imprinting is implicated in sex determination in *Nasonia*, possibly mediated by regulation of *transformer* expression (Verhulst et al. 2010). Furthermore, a recent study of the human genome suggested that gene body methylation is related to regulation of alternative transcription (Maunakea et al. 2010). It is interesting to note that even though the exon 2 is alternatively spliced between the two sexes, it is methylated in both sexes. Thus, DNA methylation may play a role to “mark” alternatively spliced exons, while not directly participating in alternative splicing per se.

In invertebrates, DNA methylation does not occur uniformly in all genes. In most invertebrate species examined, genes differ greatly in their relative levels of DNA methylation (Suzuki et al. 2007; Elango et al. 2009; Walsh et al. 2010). A widely used method to detect and characterize different levels of DNA methylation is by analyzing CpG O/E values. CpG O/E is a “normalized” level of CpG dinucleotides, which is negatively related to the levels of DNA methylation. Analysis of CpG O/E has successfully captured differences in levels of DNA methylation across different genomic regions and species. For example, in the honeybee, CpG O/E values reveal

two distinctive groups, which are correctly characterized as densely and sparsely methylated genes in subsequent experimental studies (Feng et al. 2010; Lyko et al. 2010; Zemach et al. 2010). The analysis of CpG O/E reveals a complex pattern of DNA methylation in *Nasonia*. Statistical analyses indicate that this pattern can be deciphered as a mixture of two distributions. Comparing levels of CpG depletion between *Nasonia* and the honeybee discovers several interesting evolutionary patterns. First, the low- and high-CpG genes of the honeybee and *Nasonia* exhibit functional similarity: The two groups of genes can be broadly categorized as housekeeping and development, respectively (Elango et al. 2009). Second, the densely and sparsely methylated groups of genes in these two species show different levels of conservation of DNA methylation. Specifically, genes classified as low-CpG genes in both species show greater levels of conservation compared with those classified as high-CpG genes (fig. 3, table 2). Third, this asymmetry in conservation of methylation status is directly translated into sequence similarity of *Nasonia* and honeybee genes. The numbers of nonsynonymous substitutions between *Nasonia* and honeybee genes increase with CpG O/E (tables 4 and 5). In other words, the more densely methylated genes tend to be the more conserved at the sequence level.

A unique advantage of *Nasonia* is the fact that closely related outgroup sequences are available. Using these sequences, we can investigate the relationship between molecular evolution of genes and their methylation levels over relatively short (between sibling species) and long (between *Nasonia* and *Apis*) timescales. Because the *Nasonia* species are extremely closely related, harboring 1–3% synonymous sequence divergence (Werren et al. 2010), CpG O/E classes between species do not change. Thus, we can ask whether genes in distinctive CpG O/E groups differ in their molecular evolutionary rates. Interestingly, we observed that rates of nonsynonymous and synonymous substitutions increase with the CpG O/E (tables 4 and 5). In other words, densely methylated genes at the germ lines tend to be more conserved at the sequence level, and this is observed both in the exons and introns between the sibling species. The long evolutionary timescales to *Apis* do not allow such a comparison because of saturation at synonymous and intronic sites. However, using dN, we find the same pattern of higher evolutionary rate in genes between *Nasonia* and *Apis* with increased with the CpG O/E. Importantly, this relationship is independent of the confounding effects of G+C contents (table 4) and potential functional bias (table 5).

The pattern that densely methylated genes tend to be more conserved is at odds with the mutational property of DNA methylation. Since DNA methylation tends to be mutagenetic, genes with high methylation would be expected to show higher divergence at the sequence level. The observed trends are in contrast to this expectation. The fact that we observe between the sibling species a strong correlation between CpG O/E and synonymous rates as well as between CpG O/E and intronic rates (table 4) indicates that mutation rates are likely to increase with decreasing levels of DNA methylation. The reasons for such an effect are

unclear but suggest that associations of methylation with protection of DNA (e.g., DNA packaging) or more efficient DNA repair mechanisms could be an explanation. Another possible explanation is the fact that densely methylated genes harbor fewer CpG sites than others; sparsely methylated genes harbor greater number of CpG sites. This harboring more CpG sites, even though they are only occasionally methylated, may increase the chance of mutation. An alternative (but not mutually exclusive) explanation is that highly methylated genes may have already undergone elevated substitutions at those CpG sites within the gene that can degrade in CpG without negative fitness consequences, and hence, elevated mutation rates are not expected because these substitutions occurred long ago. The apparent conservation of methylation patterns between *Apis* and *Nasonia* (table 2, fig. 3) is consistent with this idea. However, it does not explain why genes with higher CpG O/E actually have higher substitution rates on the short evolutionary timescales between the *Nasonia* species.

Our study is the first one to comprehensively analyze the relationship between levels of CpG depletion and sequence evolution. The only study that has touched upon sequence conservation in light of DNA methylation was Hunt et al. (2010), which compared the pea aphid and the honeybee, two species even more diverged than *Nasonia* and the honeybee (honeybee and the pea aphid diverged ca. 300 Ma). Due to this long divergence, they could not investigate sequence evolution in detail (e.g., they used DNA sequence identity and nonsynonymous rates, Hunt et al. 2010). Although they found that DNA sequence identity between the honeybee and the pea aphid is greater for densely methylated genes, they observed an opposite pattern using nonsynonymous rates. They found that nonsynonymous rates were greater for genes with stronger CpG depletion (Hunt et al. 2010), which is the opposite to the pattern we have observed. However, the trend reported in Hunt et al. (2010) was only weakly significant, and the reliability of estimating rates of nonsynonymous substitution in such a long evolutionary timescale is questionable. Therefore, our data on rates between closely related *Nasonia* species are particularly useful because it establishes that the pattern of lower substitution rates in methylated genes holds up over both short (dN , dS , and dI) and long (dN) evolutionary timescales. In contrast to rates of synonymous and nonsynonymous substitutions, selective constraints, measured as the ratio of nonsynonymous to synonymous substitutions (ω), do not exhibit a clear trend in relation to DNA methylation among the closely related species (table 4). Analyses of DNA sequence evolution in relation to levels of DNA methylation thus provide intriguing trends that need to be followed up in the future studies.

We also observed that similar to the pattern in the honeybee, gene lengths tend to increase with decreasing levels of germ line methylation in *Nasonia* (Results). Even though the draft sequences of two sibling species of *N. vitripennis* are available, they are not suitable for in-depth analyses of insertions and deletions at the sequence level for technical reasons. The mostly short (~ 45 bp) reads for the sibling species were aligned to *N. vitripennis*, making reliable scoring of SNPs useful

but insertion/deletions problematic (Werren et al. 2010). As correct alignment of short reads to the reference is based on sequence conservation, mapping at regions containing small insertions and deletions presents a considerable challenge to the current generation of short read assemblers. For this reason, we chose not to investigate the evolution of intron length in this study, however, with further resequencing efforts and the continued optimization of assembly software it is hoped this analysis will be possible in the near future.

Nasonia are very tractable laboratory organisms for genetics. Among their useful features are short generation time (2 weeks), ease of handling and rearing, interfertile species, and a number of genomic and genetic resources (Werren and Loehlin 2009). These features combined with the discovery of methylation in this insect, make it a promising model for studies of DNA methylation.

Supplementary Material

Supplementary table S1 and text are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

We thank the *Nasonia* genome analysis consortium for generating genome sequence data. We thank Brendan Hunt for valuable comments and discussions. This study is supported by National Research Foundation grants (M10500000126 and KRF-2008-313-C00086) to T. P., a National Institutes of Health grant (R24GM084917) to J. H. W., and funds from the Georgia Institute of Technology and a National Science Foundation grant (MCB-0950896) to S. V. Y.

References

- Bird A. 1980. DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res.* 8:1499–1504.
- Bird AP, Taggart MH. 1980. Variable patterns of total DNA and rDNA methylation in animals. *Nucleic Acids Res.* 8:1485–1497.
- Bird AP, Wolffe AP. 1999. Methylation-induced repression—belts, braces, and chromatin. *Cell* 99:451–454.
- Bock C, Lengauer T. 2008. Computational epigenetics. *Bioinformatics* 24:1–10.
- Bonasio R, Zhang G, Ye C, et al. (16 co-authors). 2010. Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Science*. 329:1068–1071.
- Duncan BK, Miller JH. 1980. Mutagenic deamination of cytosine residues in DNA. *Nature* 287:560–561.
- Elango N, Hunt BH, 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.
- Elango N, Kim S-H, Program NCS, Vigoda E, Yi SV. 2008. Mutations of different molecular origins exhibit contrasting patterns of regional substitution rate variation. *PLoS Comp Biol.* 4:e1000015.
- Elango N, Yi SV. 2008. DNA methylation and structural and functional bimodality of vertebrate promoters. *Mol Biol Evol.* 25:1602–1608.
- Farcas R, Schneider E, Frauenknecht K, et al. (12 co-authors). 2009. Differences in DNA methylation patterns and expression of the CCRK gene in human and nonhuman primate cortices. *Mol Biol Evol.* 26:1379–1389.

- Feng S, Cokus SJ, Zhang X, et al. (15 co-authors). 2010. Conservation and divergence of methylation patterning in plants and animals. *Proc Natl Acad Sci U S A*. 107:8689–8694.
- Field LM, Lyko F, Mandrioll M, Pranter G. 2004. DNA methylation in insects. *Insect Mol Biol*. 13:109–115.
- Foret S, Kucharski R, Pittelkow Y, Lockett G, Maleszka R. 2009. Epigenetic regulation of the honey bee transcriptome: unraveling the nature of methylated genes. *BMC Genomics*. 10:472.
- Fraley C, Raftery AE. 2003. Enhanced software for model-based clustering, density estimation, and discriminant analysis: MCLUST. *J Classification*, 20:263–286.
- Gavery MR, Roberts SB. 2010. DNA methylation patterns provide insight into epigenetic regulation in the Pacific oyster (*Crassostrea gigas*). *BMC Genomics*. 11:483.
- Hasselmann M, Gempe T, Schiott M, Nunes-Silva CG, Otte M, Beye M. 2008. Evidence for the evolutionary nascence of a novel sex determination pathway in honeybees. *Nature* 454:519–522.
- Hunt BG, Brisson JA, Yi SV, Goodisman MAD. 2010. Functional conservation of DNA methylation in the pea aphid and the honeybee. *Genome Biol Evol*. 2:719–728.
- Kim S-H, Yi S. 2007. Understanding relationship between sequence and functional evolution in yeast proteins. *Genetica* 131:151–156.
- Lister R, Ecker JR. 2009. Finding the fifth base: genome-wide sequencing of cytosine methylation. *Genome Res*. 19:959–966.
- Lyko F, Foret S, Wolf S, Falckenhayn C, Maleszka R. 2010. The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS Biol*. 8:e1000506.
- Maunakea AK, Nagarajan RP, Bilenky M, et al. (26 co-authors). 2010. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature*. 466:253–257.
- Raychoudhury R, Baldo L, Oliveira DCSG, Werren JH. 2009. Modes of acquisition of Wolbachia: horizontal transfer, hybrid introgression, and codivergence in the *Nasonia* species complex. *Evolution* 63:165–183.
- Regev A, Lamb MJ, Jablonka E. 1998. The role of DNA methylation in invertebrates: developmental regulation or genome defense? *Mol Biol Evol*. 15:880–891.
- Smith CR, Smith CD, Robertson HM, et al. (45 co-authors). 2011. Draft genome of the red harvester ant *Pogonomyrmex barbatus*. *Proc Natl Acad Sci U S A*. 108:5667–5672.
- Sosnowski BA, Belote JM, McKeown M. 1989. Sex-specific alternative splicing of RNA from the transformer gene results from sequence-dependent splice site blockage. *Cell* 58:449–459.
- Suzuki MM, Bird A. 2008. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet*. 9:465–476.
- Suzuki MM, Kerr ARW, De Sousa D, Bird A. 2007. CpG methylation is targeted to transcription units in an invertebrate genome. *Genome Res*. 17:625–631.
- Verhulst EC, Beukeboom LW, van de Zande L. 2010. Maternal control of haplodiploid sex determination in the wasp *Nasonia*. *Science* 328:620–623.
- Walsh TK, Brisson JA, Robertson HM, Gordon K, Jaubert-Possamal S, Tagu D, Edwards OR. 2010. A functional DNA methylation system in the pea aphid *Acyrtosiphon pisum*. *Insect Mol Biol*. 19(Suppl 2):215–228.
- Wang Y, Jorda M, Jones PL, Maleszka R, Ling X, Robertson HM, Mizzen CA, Peinado MA, Robinson GE. 2006. Functional CpG methylation system in a social insect. *Science* 314:645–647.
- Wang Y, Leung FCC. 2009. In silico prediction of two classes of honeybee genes with CpG deficiency or CpG enrichment and sorting according to gene ontology classes. *J Mol Evol*. 68:700–705.
- Weber M, Hellmann I, Stadler MB, Ramos L, Pääbo S, Rebhan M, Schübeler D. 2007. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet*. 39:457–466.
- Werren JH, Loehlin DW. 2009. The parasitoid wasp *Nasonia*: an emerging model system with haploid male genetics. *Cold Spring Harb Protoc*. 2009:pdb.emo134.
- Werren JH, Richards S, Desjardins CA, Niehuis O, Gadau J, Colbourne JK. The *Nasonia* Genome Working Group. 2010. Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science* 327:343–348.
- Wurm Y, Wang J, Riba-Grognuz O, et al. (38 co-authors). 2011. The genome of the fire ant *Solenopsis invicta*. *Proc Natl Acad Sci U S A*. 108:5679–5684.
- Xiang H, Zhu J, Chen Q, et al. (33 co-authors). 2010. Single base-resolution methylome of the silkworm reveals a sparse epigenomic map. *Nat Biotechnol*. 28:516–520.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24:1586–1591.
- Yi S, Ellsworth DL, Li WH. 2002. Slow molecular clocks in Old World monkeys, apes, and humans. *Mol Biol Evol*. 19:2191–2198.
- Zemach A, McDaniel IE, Silva P, Zilberman D. 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328:916–919.
- Zeng J, Yi S. 2010. DNA methylation and genome evolution in honeybee: gene length, expression, functional enrichment covary with the evolutionary signature of DNA methylation. *Genome Biol Evol*. 2:770–780.