Epigenetics 4:8, 551-556; November 16, 2009; © 2009 Landes Bioscience

Computational approaches for understanding the evolution of DNA methylation in animals

Soojin V. Yi* and Michael A.D. Goodisman
School of Biology; Georgia Institute of Technology; Atlanta, GA USA

DNA methylation is implicated in modulating gene transcription in a wide array of taxa. However, a complete understanding of the specific functions of DNA methylation in diverse organisms requires study of the evolutionary patterns of genome methylation. Unfortunately, investigating DNA methylation through empirical means remains challenging due to the transient nature of DNA methylation, the paucity of model systems in which to study genome methylation, and the costs associated with experimental methodology. Here we review how computational methods can be used to complement experimental approaches to further our understanding of DNA methylation in animals. For instance, comparative analyses of the molecular machinery involved in DNA methylation have been informative in revealing the dynamics of genome methylation in many organisms. In addition, analyses of specific genomic signatures of DNA methylation have furthered our understanding of the patterns of methylation within species. Finally, insight into the role of DNA methylation has resulted from computational methods used to identify specific sets of methylated genes. We suggest that an understanding of the evolution of genomic DNA methylation can be most readily achieved by integrating computational and empirical methods.

Key words: DNA methylation, comparative genomics, mutation, imprinting, phenotypic plasticity

Submitted: 10/16/09

Accepted: 10/18/09

Previously published online: www.landesbioscience.com/journals/epigenetics/article/10345

*Correspondence to:
Soojin V. Yi; Email: soojinyi@gatech.edu

Divergent Patterns of DNA Methylation in Animal Genomes

DNA methylation represents one of the most important epigenetic marks and plays an essential role in altering gene activity states.[1-5] However, genomic patterns of DNA methylation show considerable variation among taxa.[4-6] For example, vertebrate genomes are extensively methylated.[7] In contrast, invertebrate genomes display variable levels of DNA methylation.[4-10] Experimental discoveries demonstrating high variability in the patterns of DNA methylation in different animals generate many questions such as: Why does DNA methylation exist in some animals but not others? What are the functions of DNA methylation in different animal genomes? How do patterns of genomic methylation evolve in different taxa?

Unfortunately, addressing these fundamental questions remains challenging for several reasons. First, important invertebrate model organisms such as *D. melanogaster* and *C. elegans* lack DNA methylation, thereby impeding experimental work in genetically tractable systems. Second, the nature of epigenetic information means that different cells, tissues, and organisms may show divergent methylation patterns. Consequently, studies of DNA methylation are complicated by naturally occurring variation in methylation among biological entities. Finally, although considerable progress has been made in analyzing DNA methylation patterns empirically,[11,12] many techniques remain technically demanding and costly.

We suggest that computational methods represent an important counterpart to experimental means towards fully understanding the evolution of DNA methylation.[13] The purpose of this article is to discuss recent advances in understanding DNA methylation arrived at using

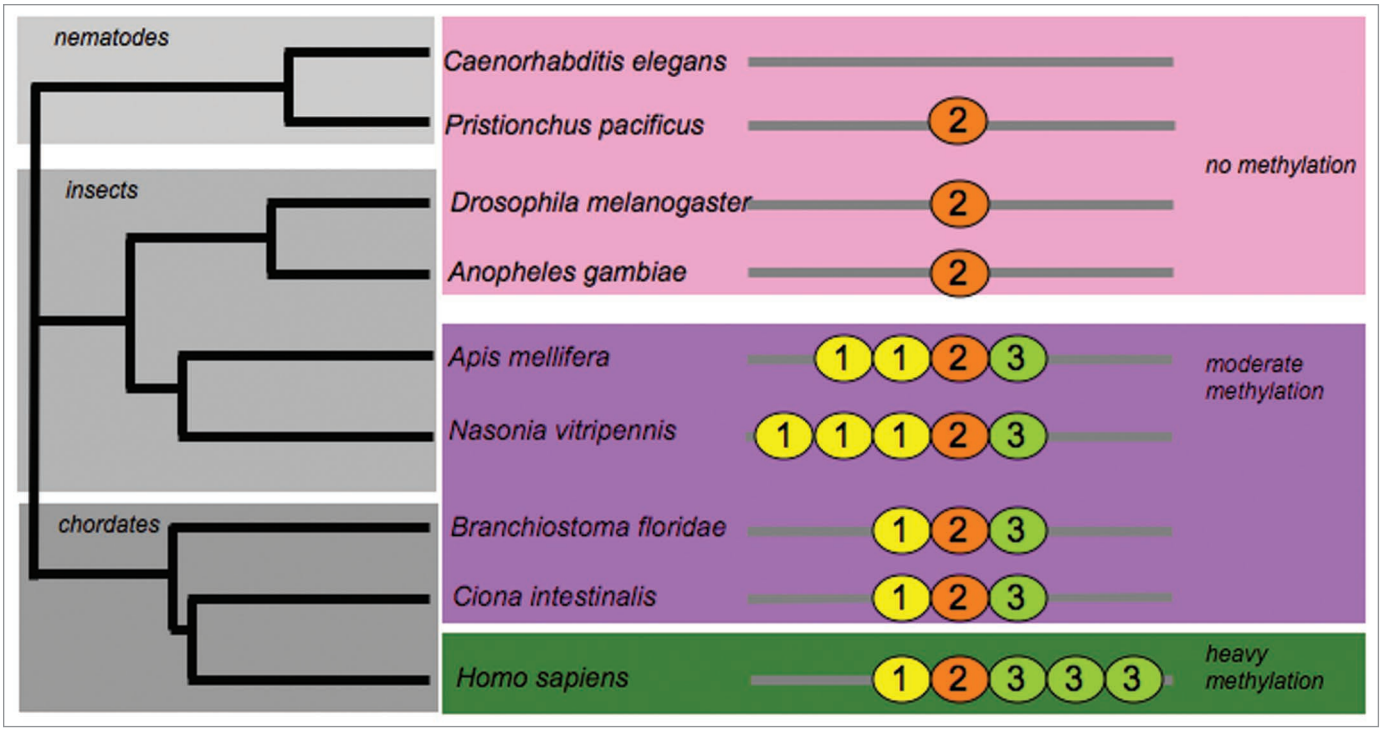

Figure 1. The numbers of different *dnmt*s in sequenced animal genomes. Different *dnmt* loci are depicted as different color circles. *dnmt2* exists in single copy and shows the most widespread distribution in animal genomes, but bears an unclear relationship with presence of functional CpG DNA methylation. In contrast, *dnmt1* and *dnmt3* families undergo frequent duplications and loss events, and seem to be associated with genome methylation.[5,17,62,63] Phylogenetic relationships among chordates are following Putnam et al.[64]

Landes Bioscience. Do not distribute.

computational analyses of genomic data. We hope to demonstrate that computational approaches can provide substantial insight into the function of DNA methylation in diverse animal taxa.

Presence of DNA Methylation Proteins Corresponds with Levels of Genomic Methylation

Determining if the proteins involved in DNA methylation processes are present within genomes, and investigating the evolution of such proteins, can provide insight into the nature of DNA methylation within species. There are at least two types of proteins directly involved in DNA methylation: those that execute the methylation process itself and those that convert information encoded in methylated genomic DNA to appropriate functional states.

DNA methyltransferases (DNMTs) perform DNA methylation and all share a conserved catalytic domain, suggesting a common and ancient origin.[14] Interestingly, the number of DNMTs in different genomes show substantial variation (**Fig. 1**).[15-17] Studies of mammalian systems have established that different DNMTs undertake distinct functions. For example, human and mouse genomes contain one DNMT1, one DNMT2, and three DNMT3s (DNMT3a/b and DNMT3L). DNMT1 is responsible for maintaining the pattern of DNA methylation between DNA replications and is referred to as the maintenance methyltransferase. DNMT3s mediate de novo methylation of previously unmethylated cytosines. The role of DNMT2 is still not completely resolved, but recent studies suggest that it may act as a tRNA methyltransferase.[18]

Comparative analyses suggest that the presence of at least one of DNMT1, DNMT2 and DNMT3 is necessary for a functional genome-wide methylation system. In addition, the numbers of different DNMTs can vary, likely leading to different patterns of genomic DNA methylation (**Fig. 1**). For example, the genome of *D. melanogaster*, which displays extremely low levels of DNA methylation, encodes only a single DNA methyltransferase, DNMT2.[6] The genome of *C. elegans* lacks DNA methylation entirely and is completely devoid of conventional DNA methyltransferases.[5] In comparison, the genome of a primitive chordate *C. intestinalis* encodes one each of the three *dnmt* genes and is known to harbor substantial DNA methylation.[8,19] Similarly, the genome of the honeybee *A. mellifera* was recently shown to encode all three DNA methyltransferase genes including two homologs of *dnmt*1, one *dnmt*2 and one *dnmt*3. As expected, functional studies confirmed the presence of an operative methylation system in honeybees.[20,21]

Complete or near complete genome sequences further suggest the existence of the complete repertoire of *dnmt*s, and, consequently, the presence of functional methylation in some insects. The genome of silkworm, *Bombyx mori*, encodes homologs of *dnmt*1 and *dnmt*2,[22] which are able to bind to purified methyl-binding domain proteins.[23] A low level of genomic DNA methylation is reported in *B. mori* and a more complete sequencing of the *B. mori* genome may reveal the presence of *dnmt*3.[2] Recently completed genomes of the wasp *Nasonia vitripennis* and the pea aphid, *Acyrthosiphon pisum*, appear

to encode the full repertoire of genes necessary for genome methylation.[17] Thus, examining the presence of different *dnmt*s from genome sequences using computational means can provide a strong clue as to whether a functional methylation system exists in a species and guide experimental characterization of lineage-specific DNA methylation.

The signal of DNA methylation conferred by DNMTs are 'read' by other proteins, including a protein family that encodes a highly conserved methyl-binding domain, hence called the 'MBD family'. MBD proteins are phylogenetically widely distributed.[15,24] Moreover, as is the case with DNMT proteins, the number of MBD proteins varies greatly between species. Genomes that exhibit heavy methylation tend to encode more *mbd*s than those with partial genomic DNA methylation.[3] Analyses of vertebrates show that *mbd*s also undergo frequent gene duplication and loss events during evolutionary history.[3,15,25] Resolving whether the presence of multiple *mbd*s reflects additional functions of DNA methylation or simply reflect functional redundancy will be critical in understanding the evolution of DNA methylation.

While it is time-consuming to elucidate functions of additional DNMT and MBD proteins experimentally, molecular evolutionary analyses of *dnmt* and *mbd* genes from related species may readily provide clues to their roles. A particularly well-established method is contrasting evolutionary rates of sites that change amino acids ('nonsynonymous sites') and sites that do not ('synonymous sites'). Such information can be used to determine if a protein's function has been optimized, if its structure is under strong conservation, or if it is adapting to a new function. This method has been very useful in identifying proteins that display lineage-specific evolution.[26-28] Analyses of DMNT and MBD protein sequences will reveal whether some methylation proteins may have been subject to rapid evolution associated with new functions. Such proteins, and their associated DNA methylation systems, should be targeted for experimental or further comparative genomic analysis because they are likely to show novel or divergent properties.

Genomic Signatures of DNA Methylation Reveal Levels and Patterns of DNA Methylation

A great deal of information concerning the nature of DNA methylation can be garnered by studying dinucleotide frequencies within genomes. DNA methylation in animal genomes occurs almost exclusively at cytosines followed by guanines, so-called 'CpG dinucleotides' (5'-CpG-3'). Methylated cytosines are chemically unstable and tend to undergo spontaneous deamination and mutate to thymines.[29,30] Thus, DNA methylation effectively leads to a high frequency of point mutations from CpGs to TpGs, and methylated genomic regions gradually lose their CpG dinucleotides. If such mutations occur in the germline, the paucity of CpGs will be inherited in the next generation, leaving signatures at the sequence level. One can then infer historical levels of germline DNA methylation from the relative CpG content encoded in DNA sequences. Moreover, even though these signatures reflect methylation-originated mutational processes in germline DNA, they have been found to be a good proxy for the levels of methylation measured in samples from different developmental stages and tissues.[13]

The depletion of CpG dinucleotides in a genomic region is often measured by the observed CpG frequency normalized by the expected CpG frequency, a metric termed 'CpG O/E'. Comparing CpG O/E values of different genomes is a straightforward, robust means to estimate levels of DNA methylation.[13] For example, Bird[31] analyzed sequence data available three decades ago and showed that CpG O/E values of vertebrate genomes were in the range of 0.2~0.4 (i.e., they harbored only 20~40% of the expected CpG dinucleotides frequencies). He further showed that several invertebrate species exhibited moderate depletions of CpG content (approximately 50~80% of expected frequencies). In comparison, the genome of *D. melanogaster* showed CpG content similar to that expected under random chance (CpG O/E ~1). Importantly, experimentally determined levels of DNA methylation and CpG O/E values showed very strong accord. For instance, Bird[32] demonstrated that vertebrate genomes exhibited 'heavy methylation', invertebrate genomes displayed 'partial methylation' and the *D. melanogaster* genome showed 'indetectable methylation'. Subsequent analyses of genomic levels of DNA methylation using modern techniques have repeatedly shown that these early determinations hold.[8,19,33]

Using the same principle, different regions of animal genomes can be scanned to determine whether DNA methylation occurs homogeneously within the genome. For example, the depletion of CpG content in the human genome is not uniform: tens of thousands of genomic regions of several hundred base pairs harbor elevated numbers of CpG dinucleotides (CpG O/E ~0.8). These 'CpG islands', which encompass only 1~2% of the genome, are maintained because they are hypomethylated in the germlines, and therefore not subject to mutations that deplete CpG dinucleotides. CpG islands can be found anywhere in the genome, but many of them are associated with promoters of broadly-expressed housekeeping genes.[34,35] Much computational effort has been put into mapping the locations of CpG islands within mammalian genomes.[13,36-39] These recent studies have greatly expanded our understanding of the characteristics of regions that undergo differential methylation in mammals.

Scanning variation of CpG O/E values within the genome has led to the discovery of divergent methylation patterns across animals. For example, Suzuki et al.[19] analyzed the distribution of CpG O/E values in *C. instestinalis*. Their analyses, using partially computational methods, revealed that a subset of transcription units were the primary targets of DNA methylation (**Fig. 2**). This pattern of 'partial' genomic methylation targeting 'gene-bodies' differs from the 'global' DNA methylation observed in vertebrate genomes, demonstrating that the pattern of genomic DNA methylation changes over time. Similarly, Elango et al.[40] and Wang and Leung[41] analyzed CpG depletion in the genome of the honeybee *A. mellifera* and found that DNA methylation was targeted to a subset of transcription units (gene bodies) (**Fig. 2**). Thus genes within *A. mellifera* showed a distinct bimodal distribution of CpG O/E suggesting that some genes are hyper-

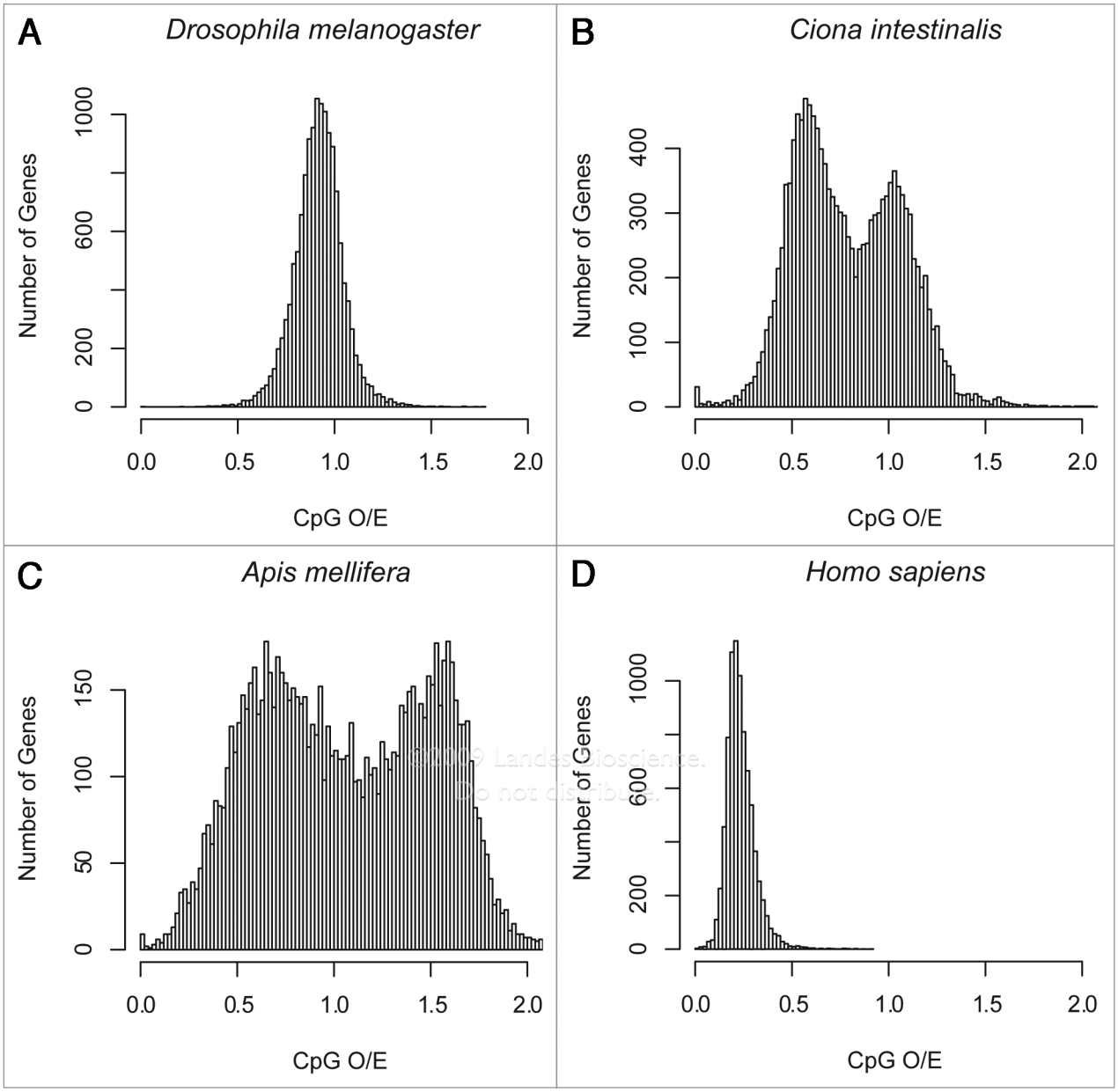

Figure 2. Distributions of CpG O/E values from different genomes indicate divergent patterns of genomic DNA methylation. (A) CpG O/E values from genes in *D. melanogaster* genome show a Gaussian distribution with the mean around 1, as expected in a system without DNA methylation. In contrast, genes in (B) *C. intestinalis* and (C) *A. mellifera* genomes exhibit distinctive 'bimodal' distributions, likely caused by the division of the genome into hyper- and hypo-methylated classes of genes. (D) CpG O/E values from human genes are around 0.2, demonstrating significant depletion of CpG dinucleotides explained by 'heavy' genomic DNA methylation.

methylated while others are hypo-methylated. Overall, comparing variation in CpG dinucleotides from different genomic regions and between species remains a powerful means for garnering information on the levels and the patterns of genomic DNA methylation.

Insights into the Roles of DNA Methylation

Elucidating the functions of DNA methylation across divergent species remains a major goal of current research.[7,42,43] Computational analyses can provide insight into the role of DNA methylation by investigating variation in putative methylation patterns among different types of genes. Here we discuss a few examples of how proposed functions of DNA methylation were, or could be, investigated using computational methods.

One of the most frequently suggested roles of DNA methylation is the control of genomic parasites such as transposable elements.[44] Evidence for this hypothesis in mammals is derived from the fact that DNA methylation appears to block mobile element transcription. Moreover, many methylated cytosines in mammalian genomes are found within mobile elements[45,46] and when DNA methyltransferases are inactivated, some endogeneous

retroviruses tend to become demethylated.[47] However, this 'genome-defense' hypothesis does not necessarily hold in other animal taxa. For example, in *C. intestinalis*, mobile elements are not methylated.[48] Computational analyses further shed doubt on this hypothesis. Suzuki et al.[49] found that repetitive elements, identified computationally, were found in both hyper- and hypo-methylated regions of *C. intestinalis* genome, thus not necessarily preferred targets for methylation. A similar result was obtained by Elango et al.[40] who found that methylation was not solely targeted to repetitive elements in the honeybee genome.

DNA methylation has also been suggested as playing a role in affecting developmental plasticity in some animals. For example, experimental and computational methods were used to demonstrate that DNA methylation modulated a very important instance of developmental plasticity in the social bee *A. mellfera*. Kucharski et al.[50] found that downregulation of a key DNA methyltransferase (DNMT3) in developing *A. mellifera* larvae resulted in profound changes in caste development trajectory. This result, demonstrating that DNA methylation was directly associated with the differentiation of the queen and worker castes in social insects, suggested that gene methylation may represent a key mechanism facilitating the evolution of social systems.[51] Elango et al.[40] used computational approaches to further understand which genes were hyper-methylated or hypo-methylated in the germline of *A. mellifera*. They found that genes that were differentially expressed between the *A. mellifera* queen and worker castes tended to have high CpG O/E values,[40] supporting the experimental work suggesting that DNA methylation was associated with the evolution of the caste system and phenotypic plasticity in social insects.[50] Thus analysis of the types of genes showing variation in CpG O/E values was informative about regulation of developmental processes.

One of the most remarkable suggested functions of DNA methylation concerns gene imprinting. Imprinting is an epigenetic phenomenon typically mediated by DNA methylation whereby the expression of an allele is influenced by the parent from which it is derived.[52] Gene imprinting is not common, but has been most famously demonstrated in mammalian loci involved in offspring growth.[53] However, gene imprinting may be a more common feature than currently appreciated. Indeed, it has been suggested that hymenopteran social insects may show substantial levels of gene imprinting because of the abundant social interactions that occur within societies and the haplodiploid genetic system displayed by hymenopteran insects.[51,54-57] How can computational analyses help in determining if imprinting occurs? Theoretical studies suggest that certain genes, particularly those associated with social behaviors or interactions,[56,57] may be evolutionary targets for imprinting. Thus, future computational studies may identify such genes using, for example, Gene Ontology annotation.[58] Moreover, recent computational analyses have incorporated features of known imprinted genes in mammals to produce sets of putatively imprinted loci,[59,60] which could be used for further analysis. For instance, imprinted genes may be expected to possess low CpG O/E values because they would be predicted to be methylated in the germlines. Thus computational methods can provide information on where empirical research should be conducted in order to determine if gene imprinting is widespread.

Finally, recent empirical results have led some to conclude that the ancestral function of DNA methylation in higher eukaryotes was the suppression of spurious transcription (i.e., transcriptional interference).[43,61] Testing such a hypothesis can be greatly aided by computational methods. In particular, determining if genes of different functions or with different levels of expression show variation in methylation patterns will provide insight into the question of whether DNA methylation is involved in suppressing spurious transcription in specific sets of genes. For example, Suzuki et al.[49] used partially computational methods to investigate the types of genes that showed variation in DNA methylation in *C. intestinalis*. They found that there was an association between putative levels of DNA methylation and gene function. These results supported the idea that CpG methylation helps suppress spurious transcription initiation in certain gene classes.

In conclusion, we suggest further integration of computational methods into the study of DNA methylation. Computational techniques are relatively rapid and inexpensive. Moreover, they promise to greatly expand our understanding of the evolution of DNA methylation in diverse taxa.

Acknowledgements

This study is supported by a NSF grant (DEB 0640690) to M.G. and S.Y. We thank Brendan Hunt for comments on an earlier version of this manuscript and for drawing Figure 2.

References

1. Bird A. Perceptions of epigenetics. Nature 2007; 447:396-8.
2. Goll MG, Bestor TH. Eukaryotic cytosine methyltransferases. Annu Rev Biochem 2005; 74:481-514.
3. Hendrich B, Tweedie S. The methyl-CpG binding domain and the evolving role of DNA methylation in animals. Trends Genet 2003; 19:269-77.
4. Klose RJ, Bird AP. Genomic DNA methylation: the mark and its mediators. Trends Biochem Sci 2006; 31:89-97.
5. Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet 2008; 9:465-76.
6. Field LM, Lyko F, Mandrioli M, Pranter G. DNA methylation in insects. Insect Mol Biol 2004; 13:109-15.
7. Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet 2008; 9:465-76.
8. Tweedie S, Charlton J, Clark V, Bird A. Methylation of genomes and genes at the invertebrate-vertebrate boundary. Mol Cell Biol 1997; 17:1469-75.
9. Mandrioli M, Volpi N. The genome of the lepidopteran Mamestra brassicae has a vertebrate-like content of methyl-cytosine. Genetica 2003; 119:187-91.
10. Kronforst MR, Gilley DC, Strassmann JE, Queller DC. DNA methylation is widespread across social hymenoptera. Curr Biol 2008; 18:287-8.
11. Fraga MF, Esteller M. DNA methylation: A profile of methods and applications. Biotechniques 2002; 33:632-49.
12. Shen L, Kondo Y, Guo Y, Zhang J, Zhang L, Ahmed S, et al. Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. PLoS Genet 2007; 3:181.
13. Bock C, Lengauer T. Computational epigenetics. Bioinformatics 2008; 24:1-10.
14. Kumar S, Cheng X, Klimasauskas S, Mi S, Posfai K, Roberts RJ, et al. The DNA (cytosine-5) methyltransferases. Nuc Acids Research 1994; 22:1-10.
15. Hung M-S, Shen C-KJ. Eukaryotic methyl-CpG-binding domain proteins and chromatin modification. Eukaryotic Cell 2003; 2:841-6.
16. Colot V, Rossignol J-L. Eukaryotic DNA methylation as an evolutionary device. BioEssays 1999; 21:402-11.
17. Maleszka R. Epigenetic integration of environmental and genomic signals in honey bees. Epigenetics 2008; 3:188-92.

18. Jurkowski TP, Meusburger M, Phalke S, Helm M, Nellen W, Reuter G, et al. Human DNMT2 methylates tRNAAsp molecules using a DNA methyltransferase-like catalytic mechanism. RNA 2008; 14:1663-70.
19. Suzuki MM, Kerr ARW, De Sousa D, Bird A. CpG methylation is targeted to transcription units in an invertebrate genome. Genome Res 2007; 17:625-31.
20. Wang Y, Jorda M, Jones PL, Maleszka R, Ling X, Robertson HM, et al. Functional CpG methylation system in a social insect. Science 2006; 314:645-7.
21. Kucharski R, Maleszka J, Foret S, Maleszka R. Nutritional control of reproductive status in honeybees via DNA methylation. Science 2008; 319:1827-30.
22. Biology analysis group, Xia Q, Zhou Z, Lu C, Cheng D, Dai F, et al. A Draft Sequence for the Genome of the Domesticated Silkworm (*Bombyx mori*). Science 2004; 306:1937-40.
23. Uno T, Nomura Y, Nakamura M, Nakao A, Tajima S, Kanamaru K, et al. Expression, purification and characterization of methyl DNA binding protein from *Bombyx mori*. J Insect Sci 2005; 5:1-8.
24. Wade PA. Methyl CpG-binding proteins and transcriptional repression. BioEssays 2001; 23:1131-7.
25. Fatemi M, Wade PA. MBD family proteins: reading the epigenetic code. J Cell Sci 2006; 119:3033-7.
26. Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, Hubisz MJ, et al. A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol 2005; 3:170.
27. Nielsen R, Hellmann I, Hubisz MJ, Bustamante C, Clark AG. Recent and ongoing selection in the human genome. Nat Rev Genet 2007; 8:857-68.
28. Yang Z, Nielsen R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol Evol 2002; 19:908-17.
29. Coulondre C, Miller JH, Farabaugh PJ, Gilbert W. Molecular basis of base substitution hotspots in *Escherichia coli*. Nature 1978; 274:775-80.
30. Duncan BK, Miller JH. Mutagenic deamination of cytosine residues in DNA. Nature 1980; 287:560-1.
31. Bird AP, Taggart MH. Variable patterns of total DNA and rDNA methylation in animals. Nuc Acids Res 1980; 8:1485-97.
32. Bird A. DNA methylation and the frequency of CpG in animal DNA. Nuc Acids Res 1980; 8:1499-504.
33. Weber M, Hellmann I, Stadler MB, Ramos L, Pääbo S, Rebhan M, et al. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet 2007; 39:457-66.
34. Antequera F, Bird A. Number of CpG islands and genes in the human and mouse. Proc Natl Acad Sci USA 1993; 90:11995-9.
35. Elango N, Yi S. DNA methylation and structural and functional bimodality of vertebrate promoters. Mol Biol Evol 2008; 25:1602-8.
36. Jiang C, Han L, Su B, Li W-H, Zhao Z. Features and trend of loss of promoter-associated CpG islands in the human and mouse genomes. Mol Biol Evol 2007; 24:1991-2000.
37. Fang F, Fan S, Zhang X, Zhang MQ. Predicting methylation status of CpG islands in the human brain. Bioinformatics 2006; 22:2204-9.
38. Das R, Dimitrova N, Xuan Z, Rollins RA, Haghighi F, Edwards JR, et al. Computational prediction of methylation status in human genomic sequences. Proc Natl Acad Sci USA 2006; 103:10713-6.
39. Takai D, Jones PA. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc Natl Acad Sci USA 2002; 99:3740-5.
40. Elango N, Hunt BH, Goodisman MAD, Yi S. DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. Proc Natl Acad Sci USA 2009; 106:11206-11.
41. Wang Y, Leung FCC. 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 2009; 68:700-5.
42. Bird A. DNA methylation patterns and epigenetic memory. Genes & Development 2002; 16:6-21.
43. Mandrioli M. A new synthesis in epigenetics: towards a unified function of DNA methylation from invertebrates to vertebrates. Cell Mol Life Sci 2007; 64:2522-4.
44. Yoder JA, Walsh CP, Bestor TH. Cytosine methylation and the ecology of intragenomic parasites. Trends Genet 1997; 13:335-40.
45. Elango N, Kim S-H, Program NCS, Vigoda E, Yi S. Mutations of different molecular origins exhibit contrasting patterns of regional substitution rate variation. PLoS Comp Biol 2008; 4:1000015.
46. Meunier J, Khelifi A, Navratil V, Duret L. Homology-dependent methylation in primate repetitive DNA. Proc Natl Acad Sci USA 2005; 102:5471-6.
47. Maksakova IA, Mager DL, Reiss D. Keeping active endogenous retroviral-like elements in check: the epigenetic perspective. Cell Mol Life Sci 2008; 65:3329-47.
48. Simmen MW, Leitgeb S, Charlton J, Jones SJ, Harris BR, Clark VH, et al. Nonmethylated Transposable Elements and Methylated Genes in a Chordate Genome. Science 1999; 283:1164-7.
49. Suzuki MM, Kerr ARW, De Sousa D, Bird A. CpG methylation is targeted to transcription units in an invertebrate genome. Genome Res 2007; 17:625-31.
50. Kucharski R, Maleszka J, Foret S, Maleszka R. Nutritional control of reproductive status in honeybees via DNA methylation. Science 2008; 319:1827-30.
51. Moczek AP, Snell-Rood EC. The basis of bee-ing different: the role of gene silencing in plasticity. Evolution & Development 2008; 10:511-3.
52. Tycko B, Morison IM. Physiological functions of imprinted genes. J Cell Physiol 2002; 192:245-58.
53. Li E, Beard C, Jaenisch R. Role for DNA methylation in genomic imprinting. Nature 1993; 366:362-5.
54. Kronforst MR, Gilley DC, Strassmann JE, Queller DC. DNA methylation is widespread across social Hymenoptera. Curr Biol 2008; 18:287-8.
55. Schaefer M, Lyko F. DNA methylation with a sting: an active DNA methylation system in the honeybee. BioEssays 2007; 29:208-11.
56. Wild G, West SA. Genomic Imprinting and Sex Allocation. American Naturalist 2009; 173:1-14.
57. Queller DC. Theory of genomic imprinting conflict in social insects. BMC Evol Biol 2003; 3.
58. The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. Nat Genet 2000; 25:25-9.
59. Luedi PP, Dietrich FS, Weidman JR, Bosko JM, Jirtle RL, Hartemink AJ. Computational and experimental identification of novel human imprinted genes. Genome Res 2007; 17:1723-30.
60. Luedi PP, Hartemink AJ, Jirtle RL. Genome-wide prediction of imprinted murine genes. Genome Res 2005; 15:875-84.
61. Bird A. Gene number, noise reduction and biological complexity. Trends Genet 1995; 11:94-100.
62. Gutierrez A, Sommer RJ. Evolution of dnmt-2 and mbd-2-like genes in the free-living nematodes Pristionchus pacificus, *Caenorhabditis elegans* and *Caenorhabditis briggsae*. Nuc Acids Res 2004; 32:6388-96.
63. Albalat R. Evolution of DNA-methylation machinery: DNA methyltransferases and methyl-DNA binding proteins in the amphioxus *Branchiostoma floridae*. Develop Genes Evol 2008; 218:691-701.
64. Putnam NH, Butts T, Ferrier DEK, Furlong RF, Hellsten U, Kawashima T, et al. The amphioxus genome and the evolution of the chordate karyotype. Nature 2008; 453:1064-71.