Supporting Information

Elango et al. 10.1073/pnas.0900301106

SI Text

Analysis of $CpG_{O/E}$ Distribution in the Honeybee Genome. In the main text, we discuss the analysis of gene bodies, because they are the best-annotated regions in the honeybee genome. In this section, we present additional analyses aimed at determining the overall patterns of DNA methylation in other regions of the genome, and report that gene bodies, defined as exons plus introns, clearly show a bimodal distribution of $CpG_{O/E}$ (Fig. S2*B*).

We investigated the distribution of $CpG_{O/E}$ of the entire genome by analyzing randomly cut 1,000 base pair segments. We found that most of the genomic segments maintain high $CpG_{O/E}$ (Fig. S2*A*), as reported by the Honeybee Genome Sequencing Consortium (1). Notably, a small portion of genomic segments was found to have low $CpG_{O/E}$, suggesting the presence of hypermethylated regions. Further analyses revealed that these low-$CpG_{O/E}$ segments likely represent gene bodies, which appear to be the main targets of DNA methylation (see the main text and below).

We then investigated whether putative promoter regions show signs of DNA methylation. For this purpose, we extracted 1,000 base pairs of sequences upstream of the transcription start sites of genes. (Our qualitative results did not change when we used 500 base pairs instead of 1,000 base pairs; results not shown.) These regions also exhibit high $CpG_{O/E}$, suggesting that promoter regions are largely hypomethylated (Fig. S2*F*). This finding is similar to the observation in a distantly related invertebrate, *C. instestinalis*, in which promoters and other intergenic regions tend to be unmethylated (2, 3). Likewise, 5′ UTRs also show a unimodal distribution of $CpG_{O/E}$, suggesting that they are largely hypomethylated (Fig. S2*D*); however, the $CpG_{O/E}$ distribution of 3′ UTRs is more complex, demonstrating a possible signature of methylation in some regions (Fig. S2*E*). Finally, the $CpG_{O/E}$ levels of introns show a "bimodal" distribution similar to that found in coding sequences, suggesting that some introns are methylated as well (Fig. S2*C*). In contrast to the patterns of $CpG_{O/E}$ found in *A. mellifera*, the whole genome, as well as genes, UTRs, and introns of *D. melanogaster*, show unimodal distributions and apparently are unmethylated (Fig. S2*G–L*).

Targeting of Transposable Elements by Methylation Is Not Evident in *A. mellifera*. It has been suggested that DNA methylation may have evolved to suppress the genomic invasion of transposable elements, because methylation and subsequent transition mutations can prohibit the proliferation of transposable elements within genomes (4, 5). But a previous analysis reported the absence of methylation of the mariner elements in *A. mellifera* (6). Consequently, we examined the possibility that selective methylation of transposable elements may explain the origin of bimodality in the *A. mellifera* genome. In particular, we investigated whether the low-CpG class is hypermethylated because it harbors substantial numbers of transposable elements. If this were the case, then bimodality in normalized CpG content would distinguish genes that harbor transposable elements and undergo DNA methylation from genes that are free from transposable elements.

To test this hypothesis, we analyzed nonrepetitive portions of honeybee genes, and found significant and clear bimodality (Fig. S3). This indicates that the presence of repetitive sequences did not bias our results. To further examine the methylation potential of transposable elements in the honeybee genome, we specifically analyzed the normalized CpG content of the mariner transposable element, the only well-annotated transposable element in the honeybee genome, which generally lacks other classes of transposable elements (1). We found a much higher CpG dinucleotide content in mariner elements than in the low-CpG class (results not shown), indicating that DNA methylation does not specifically target transposable elements in *A. mellifera*. Thus, our analyses do not support the hypothesis that the primary role of DNA methylation in social insects is to suppress transposable elements.

1. The Honeybee Genome Sequencing Consortium (2006) Insights into social insects from the genome of the honeybee, *Apis mellifera*. *Nature* 443:931–949.
2. 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.
3. Elango N, Yi S (2008) DNA methylation and structural and functional bimodality of vertebrate promoters. *Mol Biol Evol* 25:1602–1608.
4. 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.
5. Yoder JA, Walsh CP, Bestor TH (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* 13:335–340.
6. Ebert PR, Hileman JP, Nguyen HT (1995) Primary sequence, copy number, and distribution of mariner transposons in the honey bee. *Insect Mol Biol* 4:69–78.

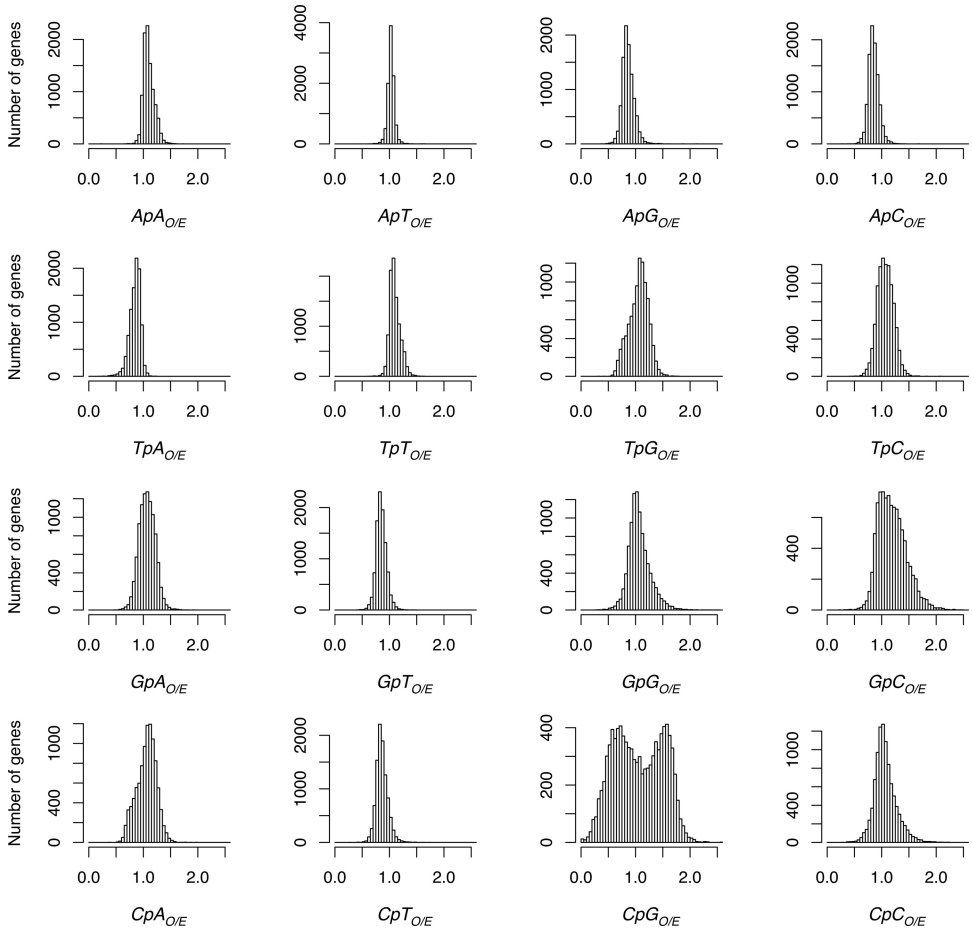

Fig. S1. Distribution of normalized dinucleotide content in *A. mellifera* genes. Only CpG$_{O/E}$ exhibits a distinct bimodal distribution, consistent with the mutational processes arising from the action of DNA methylation on CpG dinucleotides.

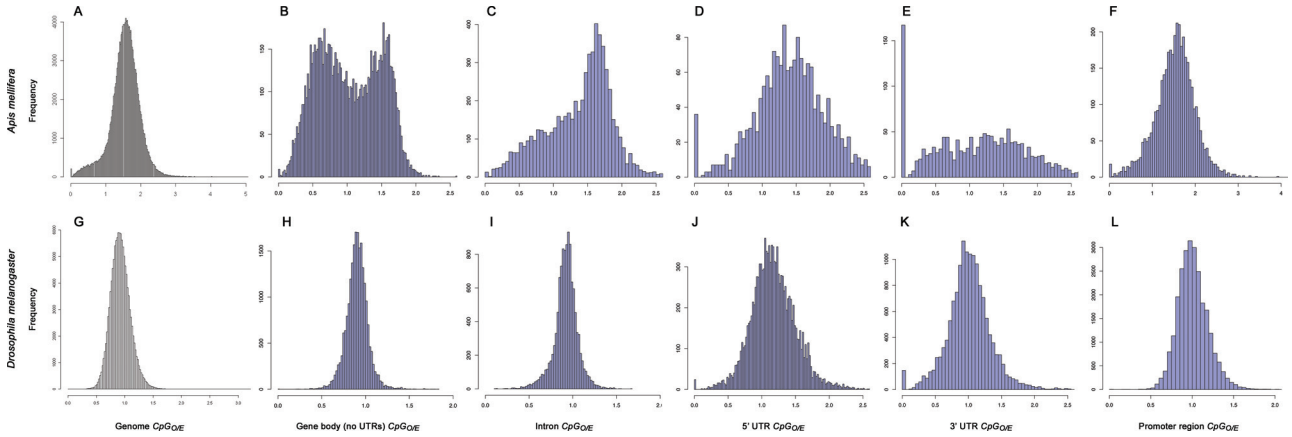

Fig. S2. Distribution of $CpG_{O/E}$ in the *A. mellifera* genome (*A*), gene bodies without UTRs (coding sequences, exons, and introns) (*B*), introns (*C*), 5′ UTRs (*D*), 3′ UTRs (*E*), and promoters (defined as 1 kb upstream of transcription start sties) (*F*) and in the *D. melanogaster* genome (*G*), gene bodies without UTRs (coding sequences, exons, and introns) (*H*), introns (*I*), 5′ UTRs (*J*), 3′ UTRs (*K*), and promoters (*L*).

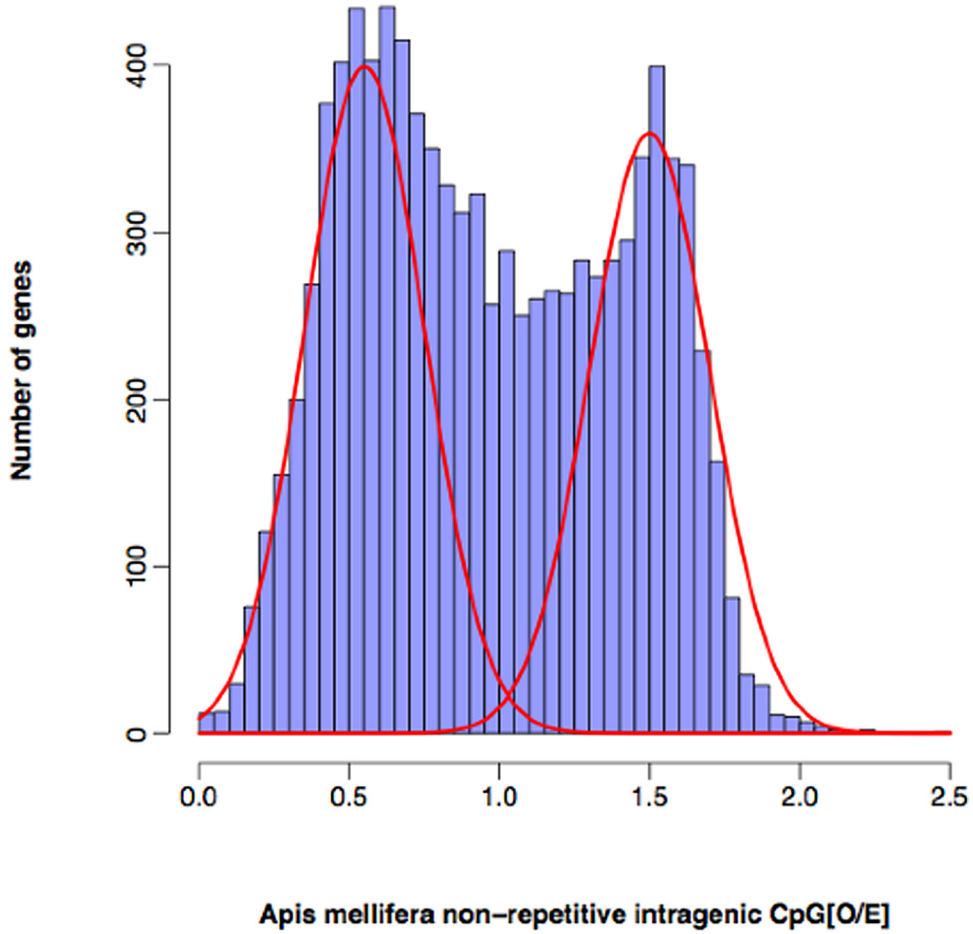

Fig. S3. Distribution of $CpG_{O/E}$ in nonrepetitive regions of gene bodies in *A. mellifera*, showing that methylation in *A. mellifera* is not localized to transposable elements.

Other Supporting Information Files

Table S1 (PDF)
Dataset S1 (PDF)
Dataset S2 (PDF)