

Full Paper

Genome-wide DNA methylomes from discrete developmental stages reveal the predominance of non-CpG methylation in *Tribolium castaneum*

Xiaowen Song^{1,†}, Fei Huang^{2,†}, Juanjuan Liu¹, Chengjun Li³,
Shanshan Gao¹, Wei Wu¹, Mengfan Zhai¹, Xiaojuan Yu¹,
Wenfeng Xiong¹, Jia Xie¹, and Bin Li^{1,*}

¹Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, Nanjing 210023, China, ²Total Genomics Solution (TGS) Institute, Bioinformatics Department, Shenzhen 518083, China, and ³Institute of Life Sciences, School of Food and Biological Engineering, Jiangsu University, Zhenjiang 212013, China

*To whom correspondence should be addressed. Tel. +86 25 85891763. Fax. +86-25-85891763. Email: libin@njnu.edu.cn

[†]The authors Xiaowen Song and Fei Huang have contributed equally to this work.

Edited by Dr. Minoru Yoshida

Received 1 November 2016; Editorial decision 25 March 2017; Accepted 29 March 2017

Abstract

Cytosine DNA methylation is a vital epigenetic regulator of eukaryotic development. Whether this epigenetic modification occurs in *Tribolium castaneum* has been controversial, its distribution pattern and functions have not been established. Here, using bisulphite sequencing (BS-Seq), we confirmed the existence of DNA methylation and described the methylation profiles of the four life stages of *T. castaneum*. In the *T. castaneum* genome, both symmetrical CpG and non-CpG methylcytosines were observed. Symmetrical CpG methylation, which was catalysed by DNMT1 and occupied a small part in *T. castaneum* methylome, was primarily enriched in gene bodies and was positively correlated with gene expression levels. Asymmetrical non-CpG methylation, which was predominant in the methylome, was strongly concentrated in intergenic regions and introns but absent from exons. Gene body methylation was negatively correlated with gene expression levels. The distribution pattern and functions of this type of methylation were similar only to the methylome of *Drosophila melanogaster*, which further supports the existence of a novel methyltransferase in the two species responsible for this type of methylation. This first life-cycle methylome of *T. castaneum* reveals a novel and unique methylation pattern, which will contribute to the further understanding of the variety and functions of DNA methylation in eukaryotes.

Key words: *Tribolium castaneum*, genome-wide, DNA methylation, life cycle, development

1. Introduction

Cytosine DNA methylation is one of the most important and conserved epigenetic modifications in diverse eukaryotic organisms.^{1–4}

This type of methylation is involved in various epigenetic processes, including X chromosome inactivation,⁵ gene imprinting⁶ and histone modification, and plays critical roles in cellular differentiation,⁷

© The Author 2017. Published by Oxford University Press on behalf of Kazusa DNA Research Institute.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Downloaded from <https://academic.oup.com/dnares/article-abstract/24/5/445/3760268>
by Queensland University of Technology user
on 12 July 2018

regulation of gene expression,^{8–10} gene alternative splicing,¹¹ suppression of transcriptional noise,^{9,12,13} and tumorigenesis.^{14–18}

In recent years, with the rapid development of next-generation sequencing technology, combining bisulphite-based detection of methylated cytosines with high-throughput whole-genome sequencing technology (bisulphite sequencing, BS-Seq) has enabled the determination of genome-wide methylomes in many eukaryotic organisms.^{2,3,19} These methylomes have significantly expanded the understanding of the distribution characteristics, functions and evolution of DNA methylation in eukaryotes.

In vertebrates, genomic cytosines are largely methylated. Most methylated cytosines occur at symmetrical CpG dinucleotides throughout the genome, except at CpG islands,²⁰ which are often observed in 5' regulatory regions, including the promoters of genes. Promoter methylation suppresses the transcription initiation of the corresponding gene and transposable elements.^{21,22} Gene body methylation is positively correlated with gene expression levels. Less non-CpG (mostly mCpA) methylation has been detected in embryonic stem cells,²³ and the functions of this type of methylation remain unclear.

Although DNA methylation is also widely observed in invertebrates, the whole-genome methylation level is much lower than that of vertebrates. Only ~0.1–0.2% of cytosines are methylated in some insects, such as the ants *Camponotus floridanus* and *Harpegnathos saltator*, the honey bee *Apis mellifera*, and the silk worm *Bombyx mori*.^{24–26} Moreover, the methylation profiles of invertebrates exhibit a mosaic pattern. Methylated cytosines are enriched in gene bodies, particularly in exons, in contrast to the continuous distribution pattern observed in vertebrates.^{26–28} The gene body methylation level is generally positively correlated with the gene expression level, while there is no relationship between promoter methylation and gene expression.²⁴ Transposons are rarely methylated in invertebrates²⁴ but are heavily methylated in vertebrates and plants.²⁹

Although the genome methylation levels, distribution patterns and biological functions of DNA methylation vary greatly among taxa, there are a number of common characteristics, including the predominant distribution of symmetrical mCG, gene body methylation, and a positive relationship between gene body methylation and gene expression levels. These features are present in most eukaryotes and were once used as criteria to judge whether an organism exhibits methylated cytosines.³⁰ A recent study demonstrated that *D. melanogaster* exhibits methylated cytosines, although the DNA methylation distribution pattern does not display the above conserved features.³¹ In *D. melanogaster*, most mCs occur at CA, CT-rich simple repetitive elements, instead of CpG dinucleotides and are enriched in intergenic regions, but absent from gene bodies. Gene body methylation is associated with lower gene expression. The most surprising finding was that DNA methylation in *D. melanogaster* is independent of DNMT2, the only known traditional DNA methyltransferase (DNMT) in *D. melanogaster*. The novel methylation pattern observed in *D. melanogaster* implies that the diversity of DNA methylation in different organisms is much more complex than previously considered and raises new questions about the origin and function of DNA methylation. Thus, characterizing the DNA methylome in additional taxa is essential for a greater understanding of the variation of DNA methylation and its biological functions in eukaryotes.

The red flour beetle *Tribolium castaneum* is an important model organism for the study of gene function. The *T. castaneum* genome encodes two DNMTs, DNMT1, and DNMT2, but lacks DNMT3, which is regarded as the only *de novo* methyltransferase and a

necessary component of the DNA methylation system.^{8,32} Therefore, the presence of DNA methylation in *T. castaneum* has long been controversial. *T. castaneum* has been considered to lack DNA methylation because the CpG(O/E) value is close to one.³³ Moreover, genome-wide bisulphite sequencing studies of *T. castaneum* in larva and adult, with average depths of 10-fold and 2.3-fold, respectively, could not detect methylated cytosines.^{3,34} However, cytosine methylation was recently detected in *T. castaneum* embryo by methylation-sensitive restriction methyltransferases and immuno-dot blot assays. Bisulphite sequencing of satellite DNA in *T. castaneum* revealed that methylated cytosines were not only restricted to CpG dinucleotides but also occurred at CpA, CpT, and CpC sites.³⁵ Until now, the true extent and pattern of genomic cytosine methylation in *T. castaneum* has remained obscure because these detection methods were either unable to localize methylation on the genome or had insufficient sensitivity to detect methylation present on only a small subset of genomic copies.

In this study, we used unbiased BS-Seq technology at an average depth of up to 40-fold to detect the DNA methylome in the four differential developmental stages of *T. castaneum*: embryo, larva, pupa, and adult. We further characterized the relationship between DNA methylation and gene expression among the different developmental stages of *T. castaneum*. Our results further expand the understanding of the function and evolution of DNA methylation in eukaryotes.

2. Materials and methods

2.1. Insect rearing

T. castaneum strain GA-1 was used in this study. The insects were reared in whole wheat flour containing 5% yeast at 30 °C.^{36,37}

2.2. Identification and cloning of *dnmt1* and *dnmt2* in *T. castaneum*

Homology searches for silkworm *dnmt* genes were conducted using the *T. castaneum* genome databases.³⁸ The initial search results were further analysed using the gene prediction software FGENESH. Based on the predicted sequences, we designed primers to amplify full-length *Tcdnmt1* and *Tcdnmt2* cDNA via PCR. Total RNA was isolated from *T. castaneum* adults using the TRIzol reagent (TAKARA, Japan). First-strand cDNA was synthesized using M-MLV Reverse Transcriptase (RNase H⁻) (TAKARA, Japan) with oligo(dT)₁₈ as the primer, and the PCR products were cloned into the pEASY-T3 cloning vector (TransGene) and sequenced by Springen (Nanjing, China).

2.3. Genomic DNA preparation and BS-Seq library generation

Genomic DNA was extracted from a pool of 50 individuals from each of the four life stages: 3 days embryo, late larva, 3 days pupa and 3 days adult, using the DNeasy Blood & Tissue Kit (Qiagen, CA). Then, 5 µg of genomic DNA from each sample spiked with 25 ng of unmethylated Lambda DNA (Promega, Madison, WI, USA) was fragmented to a size of 100–300 bp with a Covaris sonication system (Covaris, MA), followed by blunt end, 3'-end addition of dA. Cytosine-methylated adaptors were ligated to the fragmented DNA using T4 DNA ligase according to the manufacturer's instructions. Bisulphite conversion of the sample DNA was performed using a ZYMO EZ DNA Methylation-Gold kit (ZYMO Research, Irvine, CA, USA) according to the manufacturer's instructions.

The converted DNA was then amplified through nine cycles of PCR with PfuTurbo Cx Hotstart DNA Polymerase (Agilent Technologies, CA).

2.4. Illumina sequencing and alignment of methylC-seq read sequences

Paired-end sequencing was performing using Illumina Genome Analysers at a 90 bp read length. Base calling was performed using CASAVA version 1.8.2. The sodium bisulphite non-conversion rate was calculated as the percentage of cytosines sequenced at cytosine reference positions in the Lambda genome. After removing adapter sequences and low-quality sequences (N ratio over 10% or quality less than a ratio of 20 over 10%), the filtered reads were aligned to the *T. castaneum* reference genome sequences³⁹ using the general-purpose bisulphite mapping software BSMAP(v2.73) with the following parameters: -m 125 -x 234. After sorting by chromosome and position, potential PCR duplicates were removed from the aligned reads with SAMtools (v0.1.19).

2.5. Identification of mC and methylated regions (MR) in *T. castaneum*

Unconverted (methylated) cytosines are read as 'C' if the sequence is derived from the CT strand and as 'G' if the sequence is derived from the GA strand. For mC detection, only cytosine sites with a minimum coverage of four reads were used for subsequent analysis. The binomial distribution²³ was employed to correct the sequencing data to build a high-quality methylome for each of the four stages. The probability p in the binomial distribution $B(n, p)$ refers to the error rate (non-conversion and sequencing error frequency), which was calculated as the total number of sequenced Cs divided by the total sequencing depth for sites corresponding to Cs in the unmethylated Lambda genome. The bisulphite conversion rates for all samples were >99%, and the error rates were as follows: embryo, 0.0037; larva, 0.0035; pupa, 0.0039; and adult, 0.0034. n in the binomial distribution refers to the coverage depth of each potential mC. For each position with k sequenced cytosines and a total depth of n , we calculated the probability of sequencing k cytosines in n trials with an error rate of p and compared the probability of $B(k, n, p)$ to 0.01 after adjusting the P -value by the FDR. Only the mCs with adjusted P -values <0.01 were considered true positives. The methylated cytosines that passed the above processes were then used for the identification of methylated regions (MRs). First, the genomes were divided into continuous 25 bp segments, after which the segments that did not contain methylated cytosines were removed, and segments were merged into continuous segments. The resulting segments were the methylated regions.

2.6. RNA-Seq library construction and sequencing

Total RNA was extracted from each sample using the Invitrogen TRIzol® Reagent. After DNase I treatment, mRNA was enriched using oligo (dT) magnetic beads (Invitrogen). The enriched mRNA was fragmented into short fragments (~200 bp) with fragmentation buffer, and first-strand cDNA was synthesized using random hexamer primers employing the mRNA fragments as templates. Buffer, dNTPs, RNase H, and DNA polymerase I were added to synthesize the second strand. The double-stranded cDNA was subsequently purified with a QiaQuick PCR extraction kit and washed with EB buffer for end repair and dA addition. Finally, sequencing adapters were ligated to the fragments. The fragments were then purified via

agarose gel electrophoresis and enriched via PCR amplification. After a series of detection processes, the cDNA library products were sequenced using an Illumina Genome Analyser.

2.7. Mapping of RNA-Seq reads

The libraries were sequenced at the Novogene Bioinformatics Institute (Beijing, China) on the Illumina HiSeq 2500 platform and 49 bp single-end reads were generated. After base calling and filtering low-quality reads and adapters, the sequenced reads were aligned to the *T. castaneum* genome with SOAP2, allowing up to two mismatches. Reads aligned to multiple positions were discarded. Gene expression levels were calculated in RPKM (reads per kilobase of transcript per million reads).

2.8. Identification of differentially methylated regions (DMRs)

DMRs were identified in six pairwise comparisons between the different developmental stages using a 2 kb sliding window with a 1 kb step size. The number of mCs in each sliding window was counted. For valid comparison of the DMRs, the number of mCs in each sliding window was at least two. We calculated the number of differential mCs between two stages for each window. The number of differential mCs was normally distributed. The mean and SD of the differences were calculated. Regions showing a >5-fold SD difference in the global methylation level from the mean were identified as DMRs. We annotated each DMR by mapping the position of its centre relative to the features of proximal genes. TSS DMRs were centred 2 kb upstream of the TSS; TTS DMRs were centred 2 kb downstream of the TTS; and intragenic DMRs were centred from the TSS to TTS. All other DMRs were annotated as intergenic. We then performed WEGO and pathway analysis for the differentially methylated genes. The calculated P -value was corrected via the Bonferroni method.

2.9. Location of symmetrical mCG sites in the genome of *T. castaneum*

We located symmetrical mCG sites in the *T. castaneum* genome of the four developmental stages. The TSS region was 2 kb upstream of the TSS; the TTS region was 2 kb downstream of the TTS; and the intragenic region was the region from the TSS to the TTS. All other regions were defined as intergenic regions.

2.10. Validation of DNA methylation

Two methods were applied to validate the methylation status in *T. castaneum*. First, global DNA methylation levels in the four stages of *T. castaneum* were determined using a methylated DNA quantification kit (Epigentek, P-1034, USA) according to the manufacturer's protocols, and specific methylated sites were validated through methylation-sensitive restriction enzyme-mediated PCR. Three methylation-sensitive restriction enzymes, *MspI* (NEB, R0106V, USA), *HpaII* (NEB, R0171V, USA), and *ClaI* (NEB, R0197V, USA) were selected for the validation procedure. *MspI* and *HpaII* are isoschizomers; both enzymes recognize the sequence 5'-CCGG-3' and exhibit differential methylation sensitivity. *MspI* cannot cleave the CCGG sequence when the external C is methylated, while *HpaII* is blocked when either the external or internal cytosine is methylated. For *ClaI*, which recognizes the sequence 5'-ATCGAT-3', the enzyme is inhibited when the cytosine in the recognition sequence is methylated.

Genomic DNA was extracted from the four life stages (3 d embryo, late larva, 3 days pupa and 3 days adults) using the DNeasy Blood & Tissue Kit (Qiagen, CA). Then, the digestion reaction was performed using genomic DNA (1 µg) and 1 µl of the enzymes. The reactions were incubated at 37°C overnight to ensure complete digestion. Several methylated regions containing the enzyme recognition sequences were randomly selected and amplified according to the sequencing results. For these assays, the primers were designed to encompass a single recognition site in the corresponding PCR products (Supplementary Table S7). The amplification reactions were performed using the above digested DNA as a template. All experiments were carried out with three biological replications.

3. Results

3.1. DNA methyltransferases in *T. castaneum*

Cytosine DNA methylation in eukaryotes is catalysed by a family of DNA methyltransferases (DNMTs), which are classified into three groups (DNMT1, DNMT2, and DNMT3) according to their functional mechanism and substrate preference.^{8,32} A homology search against the *T. castaneum* genome identified one orthologue each for *dnmt1* and *dnmt2*. There was no homologous sequence for *dnmt3*, consistent with previous reports.³⁵

To determine whether the predicted sequences of these two candidate DNMT genes were correct, the full-length cDNAs of *dnmt1* and *dnmt2* were obtained via RT-PCR. Comparison of the sequence of the putative full-length *Tcdnmt2* clone (KX553975) with the

predicted *dnmt2* sequence indicated that it was composed of three exons capable of encoding a protein with 329 amino acids (Fig. 1B).

However, a comparison of the *Tcdnmt1* cDNA clone sequence with the predicted coding sequence (CDS) at Beetlebase revealed the addition of a 140 bp sequence between exons 8 and 9 in the cDNA clone sequences. We designated the 'superfluous exon' exon 8a. Inspection of the 'superfluous exon' sequence indicated that it may be an intron because it begins with a 5'GT and ends with a 3'AG, which are typical characteristics of introns. To validate the reliability of the clone sequence, we designed a pair of primers flanking exon 8a and amplified the target sequences. We obtained two cDNA fragments differing in size by ~140 bp (Supplementary Fig. S1), which revealed that there were two different *Tcdnmt1* transcripts. We next used the same primers to screen out and sequence the two alternatively spliced forms of *Tcdnmt1* among the DNA clone recombinants. Comparison of the two alternatively spliced transcripts of *Tcdnmt1* revealed that the shorter *Tcdnmt1* sequence, *Tcdnmt1a* (KX553973), was consistent with the predicted sequences in Beetlebase, whereas the longer sequence, *Tcdnmt1b* (KX553974), included an additional 140 bp sequence (exon8a) (Fig. 1A). Comparison of the relative expression levels of the two genes revealed that *Tcdnmt1b* was predominately expressed in all life stages of *T. castaneum* (Supplementary Fig. S1). The inclusion of exon 8a led to a shift in the reading frame and shortened the ORF due to a stop codon at the end of exon 8a. Thus, *Tcdnmt1a* encodes a protein with 1,208 amino acids, while the translation product of *Tcdnmt1b* is a 729 amino acid protein. The two proteins share 682 amino acids, starting from their N-termini (Supplementary Fig. S2).

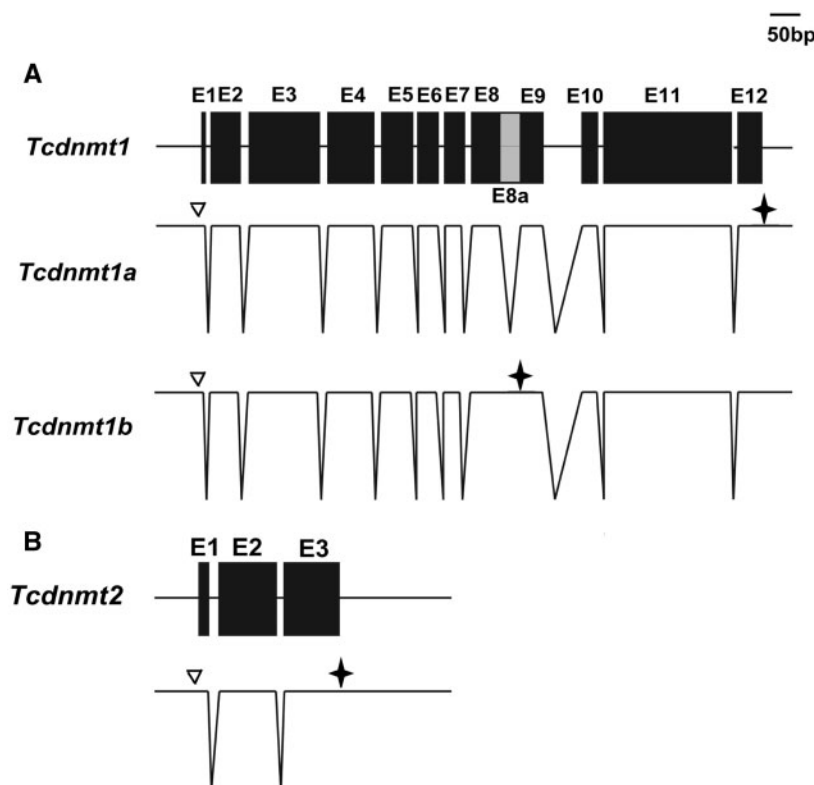


Figure 1. Schematic diagram of the exon-intron organization of the putative *T. castaneum* *dnmt* genes. (A, B) The exon-intron organization of each *dnmt* gene was determined via sequence comparison between the genomic sequence and the available cDNA sequence. (A) Alternatively spliced forms of *Tcdnmt1* (*Tcdnmt1a*, *Tcdnmt1b*). (B) Gene organization of *Tcdnmt2*. The black rectangles represent the exons (E), and lines represent the introns. The grey rectangles indicate exon8a of *Tcdnmt1b*. The open arrowheads represent the open reading frame start codon, while the asterisks indicate the stop codon.

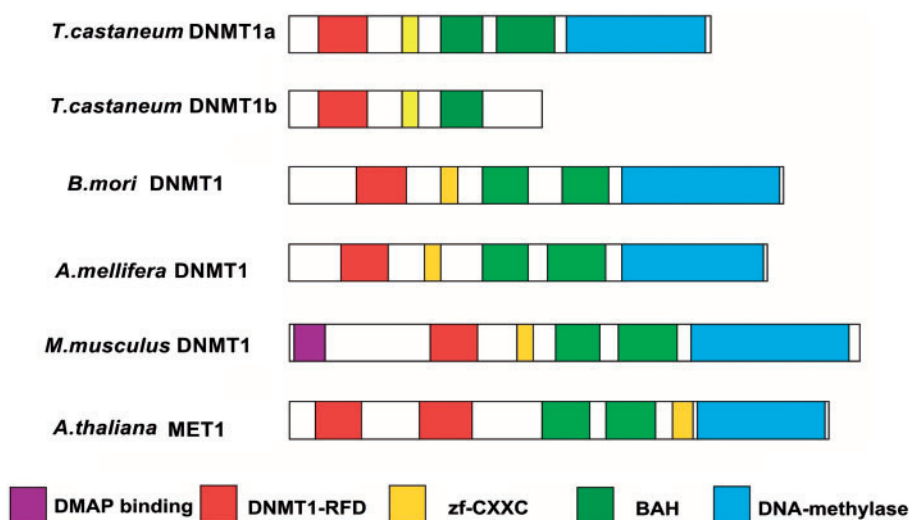


Figure 2. Conserved functional domain organization of the proteins of two different transcripts of *Tcdnmt1* and DNMT1 proteins in other organisms, including *Bombyx mori*, *Apis mellifera*, *Mus musculus* and *Arabidopsis thaliana*.

Next, we analysed the conserved functional domain organization of these two translation products of *Tcdnmt1* using SMART software. *Tcdnmt1a* shares conserved functional domains with other organisms, whereas *Tcdnmt1b* lacks the C-terminal DNA methylase domain (Fig. 2). The DNA methylase domain is the catalytic domain of DNMTs responsible for the transfer of methyl groups from SAM to the corresponding 5-C in the pyrimidine ring of the DNA. The absence of the DNA methylase domain indicates loss of the basic function of DNMTs. Thus, *Tcdnmt1b* likely cannot methylate DNA.

3.2. Detection of methylcytosines in the four life stages of *T. castaneum*

To investigate the DNA methylation pattern of *T. castaneum*, we performed Illumina whole-genome bisulphite sequencing for the four developmental stages of *T. castaneum*: 3 days embryo, late larva, 3 days pupa, and 3 days adult. We initially generated 191.7, 176.3, 178.4, and 163.2 million raw reads for each sample, respectively. After removing low-quality and clonal reads, we obtained a total of 145.6, 151.6, 141.7, and 134.3 million clean reads, respectively, and aligned the reads to the *T. castaneum* reference genome, which mapped 81.7% of the reads to the embryo, 80.1% to the larva, 87.4% to the pupa, and 86.9% to the adult. In total, the average read depth was 43.2X, 44.1X, 45.0X, and 42.4X, respectively, per strand for each sample (Supplementary Table S1). On average, more than 97% of cytosines in the *T. castaneum* genome were covered by at least one sequence read in each of the four samples (Supplementary Fig. S3). From the control Lambda DNA (Materials and Methods), we calculated false-positive rates of 0.0037, 0.0035, 0.0039, and 0.0034, respectively, for the four stages. We then applied the error rates for each stage to correct the mC identification via the method of the Lister *et al.*,²³ which is based on the bimodal distribution and false discovery rate constraints.

After correction, we identified 32,699, 75,891, 98,622, and 104,078 methylated cytosines in the embryo, larva, pupa and adult genomes, respectively, accounting for 0.08%, 0.19%, 0.24%, and 0.26%, of all genomic cytosines (Supplementary Table S2). The relatively low methylation level found in *T. castaneum* appears to be a common feature of insects and is much lower than in vertebrates.

Although the global methylation level was low, the DNA methylation level increased during the transition from the new-born to mature stages, consistent with observations in *Trichinella spiralis*.⁴⁰ In the human genome, the global methylation level is reduced as the cell differentiation level increased.²⁰ The changes in the DNA methylation level may reflect the response to both intrinsic and environmental exposure. The increasing tendency of DNA methylation in *T. castaneum* may play important roles during the normal development of this species.

3.3. Context and degree of DNA methylation in *T. castaneum*

We assessed the relative prevalence of DNA methylation in each sequence context throughout the genome in each developmental stage. The four stages exhibited similar distribution patterns: almost 75% of the methylated cytosines were in the CHH (H indicates non-G bases) context, while 13% and 12% were in the CG and CHG contexts, respectively (Fig. 3A). We further analysed the mCN (N indicates A,G,T,C bases) base composition to determine whether there was a significant base preference at the +1 position relative to cytosine. All four samples showed a strong preference for cytosine methylation in the CpA context (Fig. 3B).

Moreover, for mCG, mCHG, and mCHH, the methylation levels of most methylated sites ranged from 10% to 30%, with similar distributions among the four developmental stages (Fig. 3C). Since mCs are relatively sparse in *T. castaneum* compared with vertebrates, we identified methylated regions (MR) of the genome using relatively dense mCs. Visualization of the alignments at one randomly selected MR from the embryo sample revealed an allelic pattern of methylation in which methylation was concentrated in a subset of sequence reads (Supplementary Fig. S4), which may imply that the methylated sites present in *T. castaneum* are restricted to a small subset of cells, or that the distribution pattern is variable from cell to cell.

Moreover, the density of DNA methylation exhibited large variations across each chromosome (Fig. 3D and Supplementary Figs S5–S8), indicating a mosaic pattern in which the hypermethylated regions were interspersed among the larger, unmethylated regions.

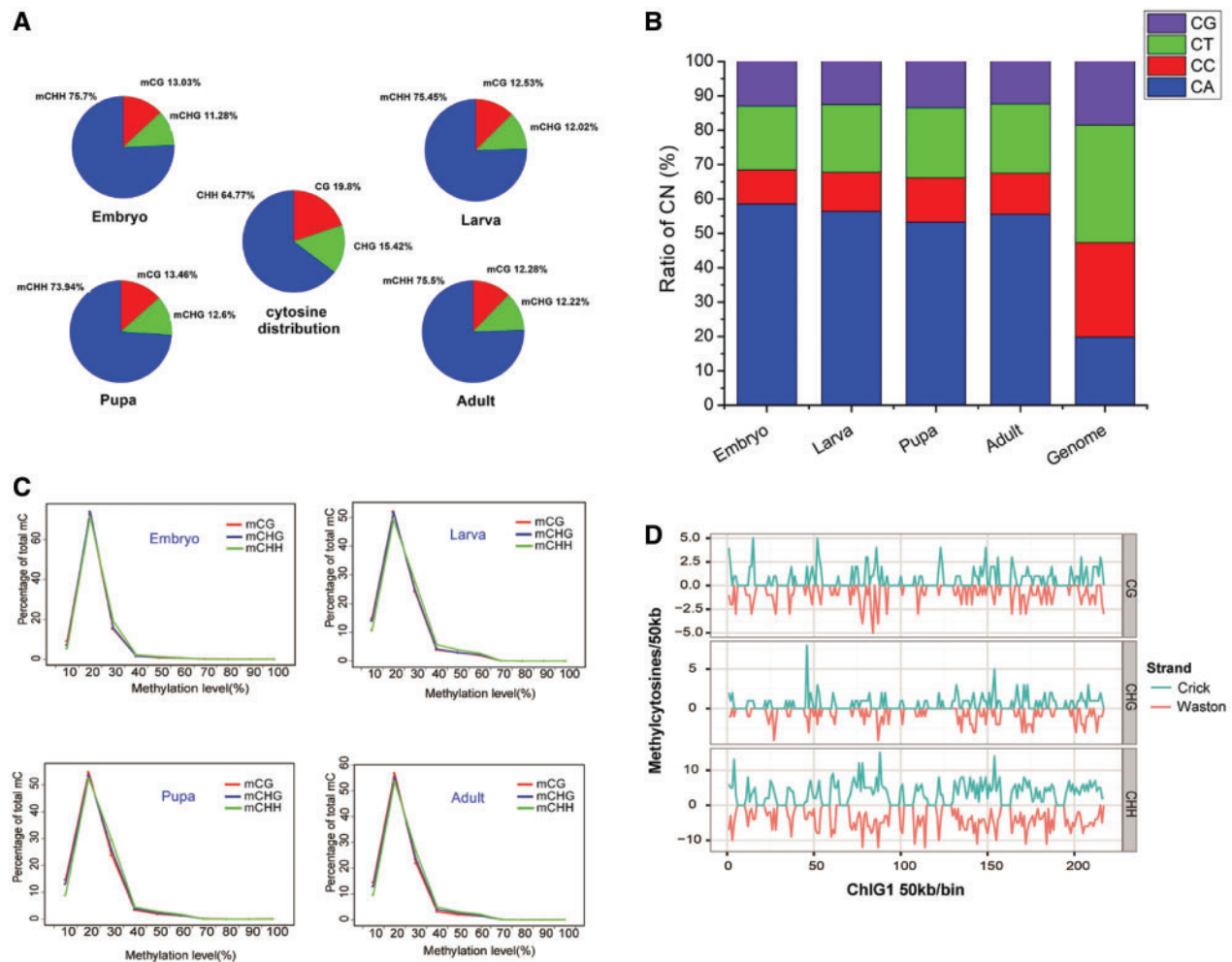


Figure 3. DNA methylation patterns and chromosomal distribution in the four stages of *T. castaneum*. (A) The fraction of mCs identified in each sequence context in the four stages in comparison with the fraction of all Cs in each sequence context in the genome. The H indicates non-G bases. (B) DNA methylation in various combinations of mCN sequence contexts (N indicates any of the four nucleotides) compared with the fraction of all Cs in the four contexts in the genome. (C) The percentage methylation level distribution of mCs. (D) The density of mCs identified on the two DNA strands of chromosome 1 in embryos.

This appears to be a common distribution pattern of DNA methylation in insects.^{2,3,41}

3.4. Methylated cytosines are absent from genes and enriched in intergenic regions

To understand the functional significance of DNA methylation in *T. castaneum*, we analysed the methylation profiles of different functional elements, including genes (coding sequences + introns), 3'UTRs, 5'UTRs, intergenic regions, and repeat elements. Both the methylation density (number of mCs divided by the total number of C sites covered in the region of interest) and percentage of methylation (fraction of the total 'C' read number within the total 'C' and 'T' read number for each genomic region) were used as predictor variables. In *T. castaneum*, compared with the average level in the genome, the methylation level was lower in CDS and higher in intergenic regions and introns (Fig. 4A and B). However, in most eukaryotes, genes (particularly the CDS regions) are heavily methylated.³ Thus, we further analysed the methylation profiles of each protein-coding gene, including the gene body and the 2 kb regions upstream

and downstream. After normalizing the methylation profiles for all genes, we observed that the gene body methylation level was lower than that of the flanking intergenic regions: similar patterns were observed among the four developmental stages (Fig. 5A and B). There was high methylation throughout the upstream region of the gene body, with a sharp dip immediately upstream of the ATG. Methylation remained at a low level in the gene body and then exhibited a sharp spike at the stop codon, with recovery to the initial methylation level being observed in the upstream regions.

3.5. Methylated cytosines accumulate at introns and exhibit a strong exon/intron pattern

We next examined the methylation profiles of exons, introns and exon–intron boundaries. Methylation was significantly higher in introns and declined in adjacent exons, revealing a strong exon/intron methylation pattern (Fig. 6). Only the distribution pattern of *D. melanogaster* is similar to that of *T. castaneum*, with methylated regions being enriched in introns and intergenic regions. In most other eukaryotes,³ methylated cytosines are concentrated in exons, whereas introns are rarely methylated. The different methylation

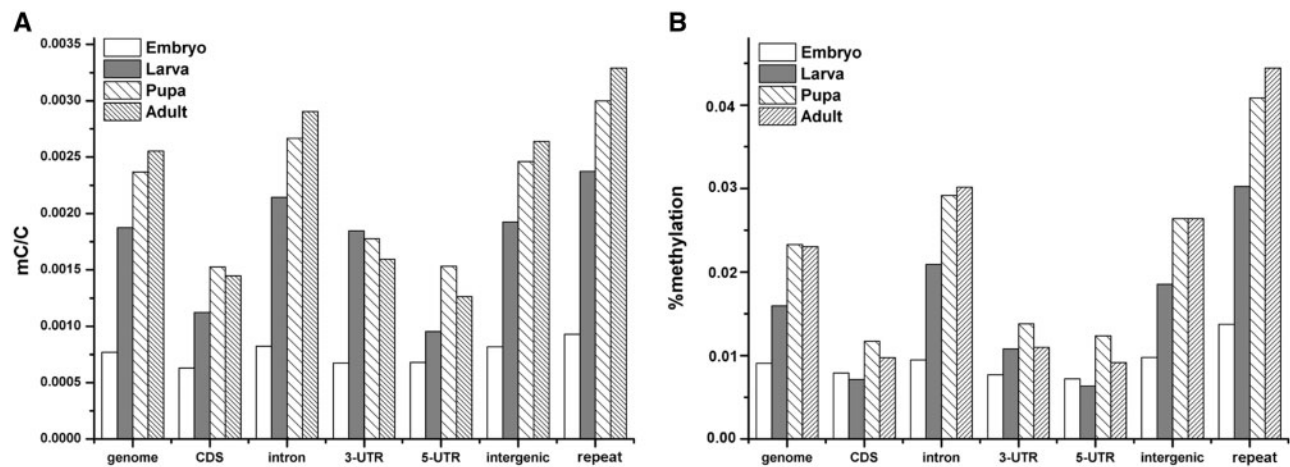


Figure 4. Methylation profiles for different functional regions in the four stages of *T. castaneum*. (A, B) Methylation density (A) and percentage methylation level (B) (y-axis) in different functional regions (x-axis).

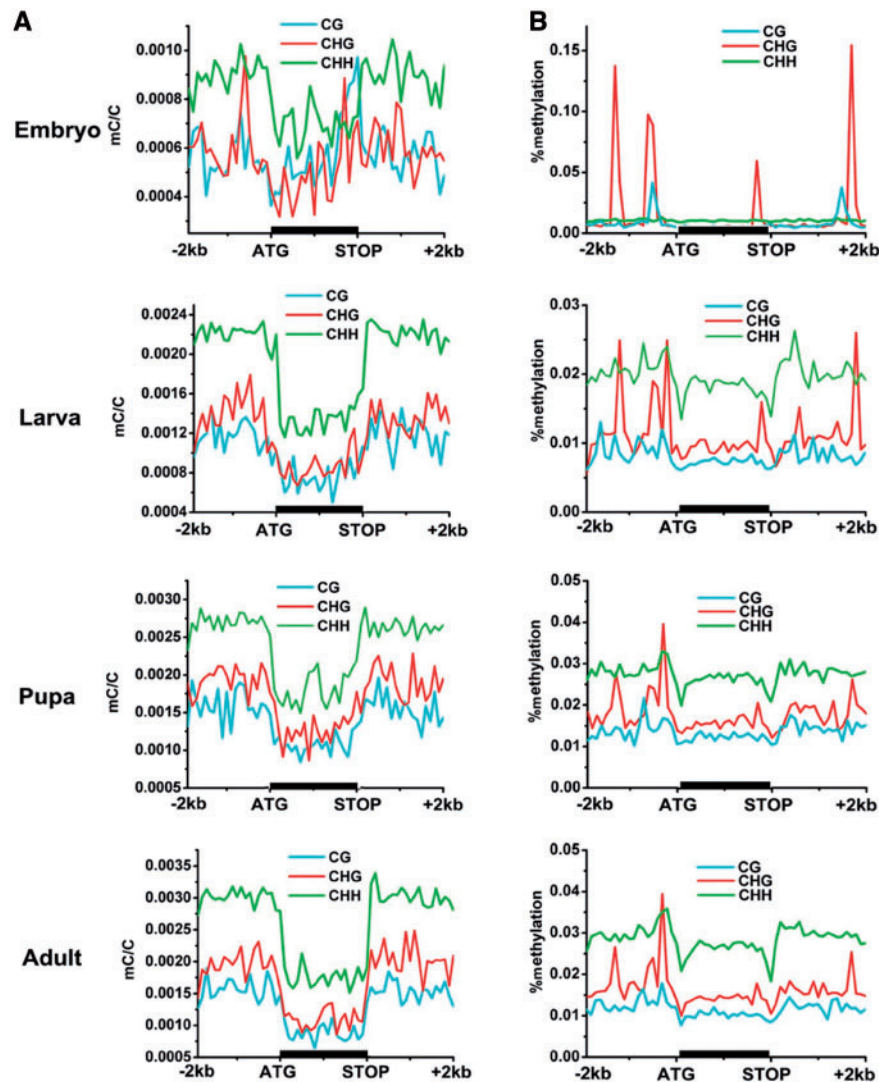


Figure 5. Methylation patterns of protein-coding genes. (A, B) Methylation density (A) and percentage methylation level (B) around genomic regions. The regions 2kb upstream and downstream of the gene and the gene body were divided into 20 intervals.

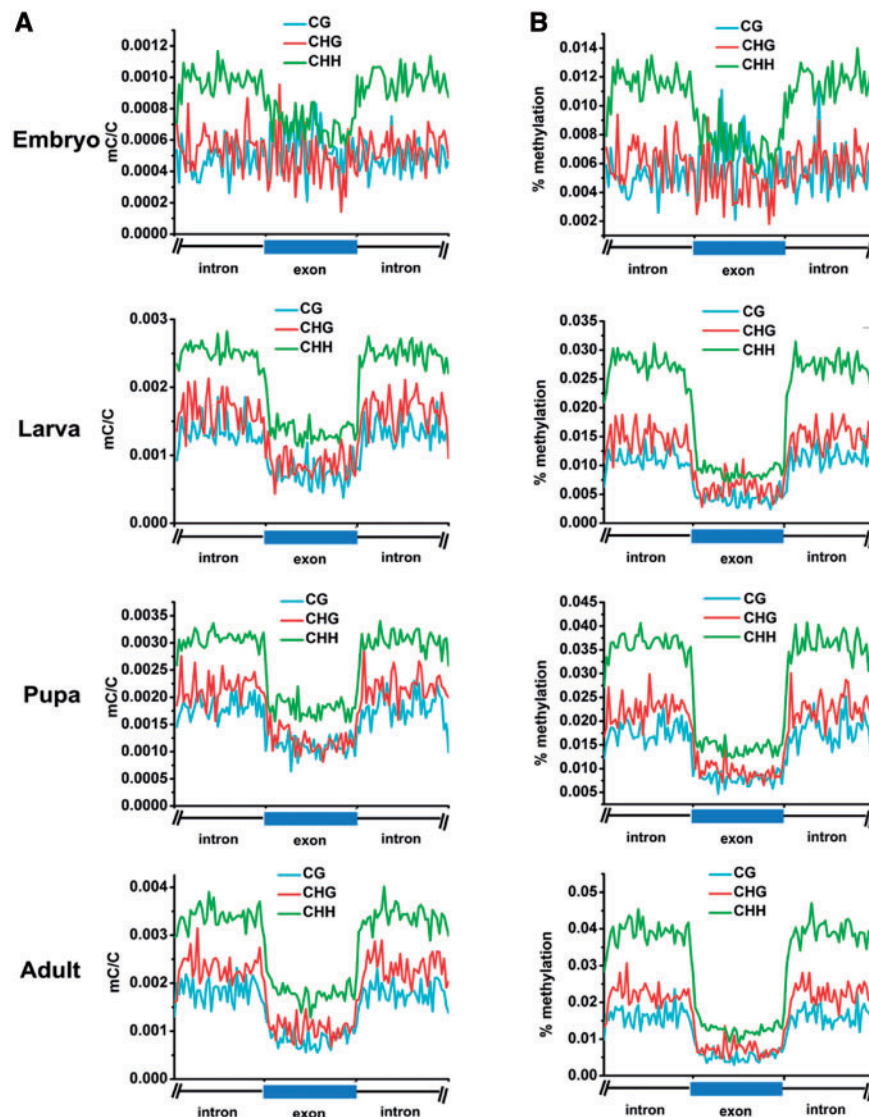


Figure 6. Methylation profiles of exons, introns and junction boundaries. (A) Methylation density of exons and introns. (B) Percentage methylation level of exons and introns.

patterns of exons and introns may indicate a link between DNA methylation and transcription or the splicing machinery.

Moreover, the methylation level was almost evenly distributed in each exon or intron of the gene body in *T. castaneum* (Supplementary Fig. S9). In *Camponotus floridanus*, *Harpegnathos saltator*, *Nasonia vitripennis*, and *Bombyx mori*,^{26,28} gene body methylcytosines are enriched in the 5' coding region and decline towards the 3' coding region, with methylation peaks typically occurring at the second exon and intron.

3.6. Repeat elements are heavily methylated

The methylation pattern of repeat elements varies greatly among species. Repeat elements are heavily methylated in plants and mammals but rarely methylated in insects. Notably, in *T. castaneum*, the average methylation level of repeat elements was higher than the average genome level (Fig. 4). We further detected the DNA methylation profiles in repeat regions and the flanking 2 kb regions, which revealed that the methylation level was significantly higher in repeat elements

than that in the flanking regions (Fig. 7). We next examined the methylation levels of different classes of repeat elements, including transposable elements, satellites, simple repeats, and other repeat elements (Supplementary Fig. S10). Among the repeat elements, the simple-repeat elements exhibited the highest methylation level, consistent with findings in *D. melanogaster*. In certain classes of transposable elements, such as LINES, the DNA was more highly methylated than the average genome level, whereas others, such as LTRs and RCs, were less methylated in *T. castaneum*.

3.7. Relationship between DNA methylation and gene expression

To determine the relationship between DNA methylation and gene expression, we measured differential gene expression in the four life stages of *T. castaneum* using Illumina high-throughput RNA-Seq technology on mRNA extracted from the four samples. Most of the raw reads (6,130,742, 5,850,925, 5,957,631, and 6,115,496 reads for 3 days embryo, late larva, 3 days pupa, and 3 days adult,

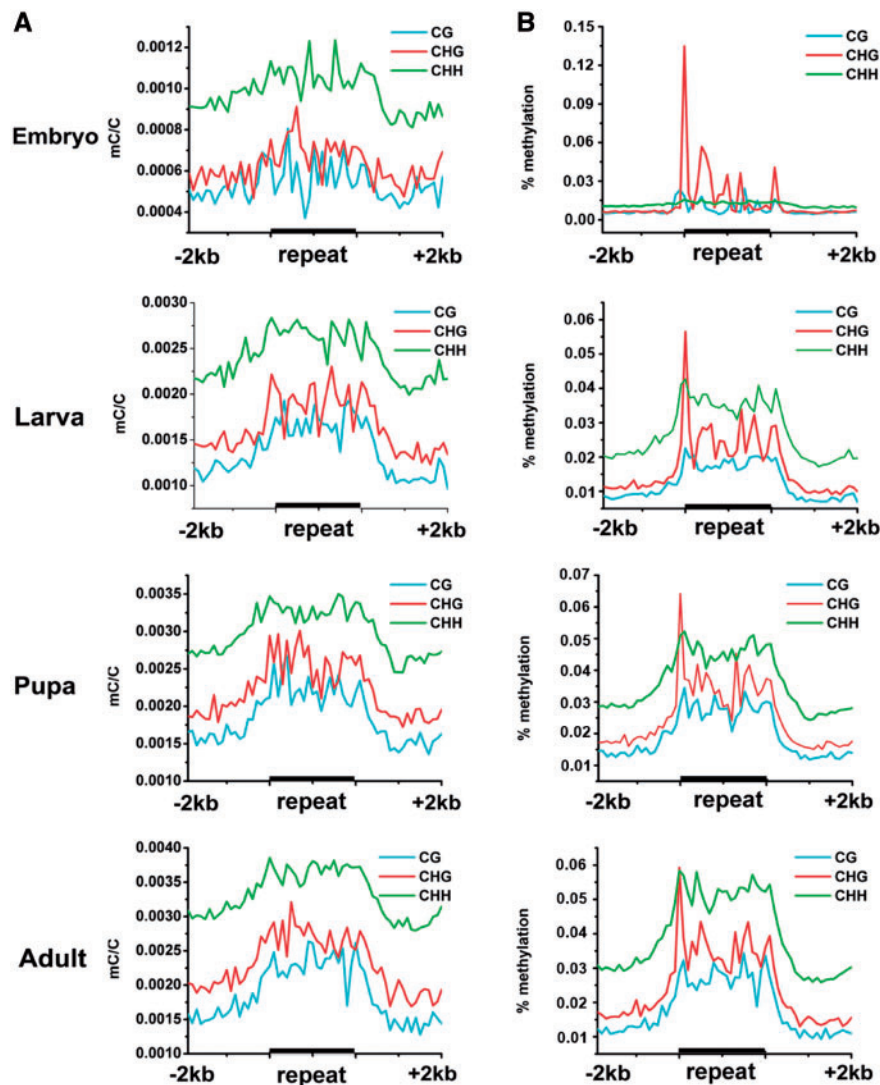


Figure 7. Methylation density and percentage methylation level around repetitive elements. Regions 2 kb upstream and downstream of repetitive elements and repetitive elements were divided into 20 intervals. The methylation level was calculated for each bin, and the average for all bins is shown.

respectively) were uniquely mapped to previously annotated genes (52.37%, 61.06%, 60.75%, and 50.23%, respectively), and there were 12,765, 12,496, 13,000, and 13,322 genes, respectively, with at least one unique read detected in the four stages. Among these genes, 11,252 genes were expressed in all four life stages, while 250 embryo-specific, 153 larva-specific, 171 pupa-specific, and 408 adult-specific expressed genes were identified (Supplementary Fig. S11).

Consistent with our previous analysis, gene regions were observed to be rarely methylated. Almost half of the genes in the *T. castaneum* genome did not contain methylated cytosines, and no more than five cytosines were methylated in most other genes. Next, we further analysed the expression levels of methylated genes (genes containing at least one methylcytosine) and unmethylated genes (genes without methylcytosine) in the four developmental stages to determine whether methylation was correlated with gene expression. We observed average expression of 72.15 RPKM for the 9,934 unmethylated genes and 56.53 RPKM for the 3,685 methylated genes in the embryo sample ($P=0.000195$, Wilcoxon two-sided test); 81.85 RPKM for the 8,266 unmethylated genes, and 53.71 RPKM for the

5,355 methylated genes in the larva sample ($P=0.009852$); 73.52 RPKM for the 7,194 unmethylated genes and 55.87 RPKM for the 6,425 methylated genes in the pupa sample ($P=4.34E-07$); and 77.18 RPKM for the 7,332 unmethylated genes and 53.87 RPKM for the 6,289 methylated genes in the adult sample ($P=0.02678$) (Table 1). Overall, the DNA methylation level of *T. castaneum* genes, although low, appeared to be negatively correlated with the gene expression level.

To further determine whether there was a linear relationship between gene expression levels and methylation levels, we measured the expression levels of genes with different methylated cytosine sites. However, we did not observe a linear correlation between the gene expression level and methylation level (Supplementary Fig. S12). Therefore, gene expression was correlated with methylation status but did not decrease with an increasing methylation level in methylated genes.

3.8. DMRs associated with differentiation

To gain insights into whether DNA methylation plays roles in the transitions between different developmental stages by regulating

Table 1. Expression levels of methylated genes and unmethylated genes in the four stages of *Tribolium castaneum*

Sample	Number of methylated genes	Average expression level (RPKM)	Number of unmethylated genes	Average expression level (RPKM)	P-value
Embryo	3685	56.52654134	9934	72.14988237	0.000195
Larva	5355	53.70656222	8266	81.85341874	0.009852
Pupa	6425	55.86642745	7194	73.52457263	4.34E-07
Adult	6289	53.86641288	7332	77.17578712	0.02678

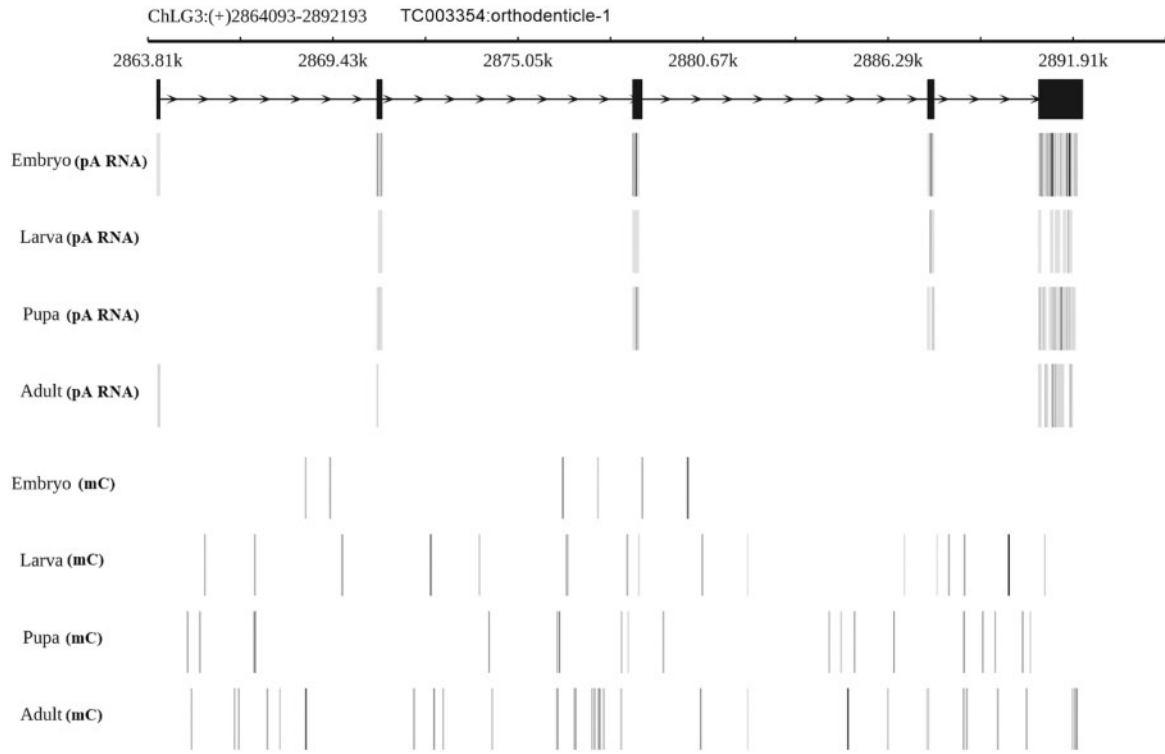


Figure 8. Differentially methylated genes. Examples of DMRs found in the homeobox transcription factor *orthodenticle-1* (TC003354), which contains a region that is hypomethylated in embryos compared with the other three stages, in which the gene is expressed at higher levels. pA RNA tracks indicate the normalized RNA-Seq signal according to the intensity of black colouring; mC tracks show methylated cytosines as vertical lines, and their degree of methylation is indicated by the intensity of black colouring (black, all reads methylated; white, all reads unmethylated).

gene expression, we identified DMRs through pairwise comparison between the different developmental stages and analysed whether the genes associated with the DMRs exhibited stage-specific expression. The overall DNA methylation patterns of the four samples were highly correlated. We identified 22 DMRs between the embryo and larva, 93 DMRs between the larva and pupa, and 115 DMRs between the pupa and adult (Supplementary Table S3). Most of the DMRs were associated with gene regions (promoter, gene body, TTS). Genes associated with transcription and development were differentially methylated. The gene *orthodenticle-1* is a gap gene encoding a homeobox transcription factor that plays a key role in insect embryogenesis. Compared with other stages, this gene exhibited significantly lower gene methylation in embryo, consistent with its higher expression and functional importance in this stage (Fig. 8). We next performed Gene Ontology (GO) analysis for the genes associated with the DMRs between different developmental stages. The genes that were differentially methylated between the embryo and larva were enriched for GO terms related to cellular and metabolic

processes, binding and catalytic activity. The differentially methylated genes identified in the larva compared with pupa were enriched for biological phase, enzyme regulator and transporter activity. The genes that were differentially methylated between the pupa and adult were significantly enriched in reproductive processes, the response to stimuli and signalling (Fig. 9, Supplementary Figs S13–S17).

3.9. Symmetrical methylated CpG sites occupy a small part of the *T. castaneum* methylome

As symmetrical methylated CpG sites predominate in the methylomes of most eukaryotes, we next asked whether this type of methylation exists in *T. castaneum*. Therefore, we analysed symmetrical mCpG sites in the methylomes of the four stages. We identified 106, 64, 98, and 98 symmetrical mCG sites in embryo, larva, pupa, and adult, respectively, which occupied 2.49%, 0.67%, 0.74%, and 0.76% of the mCG sites in the whole genome. The average

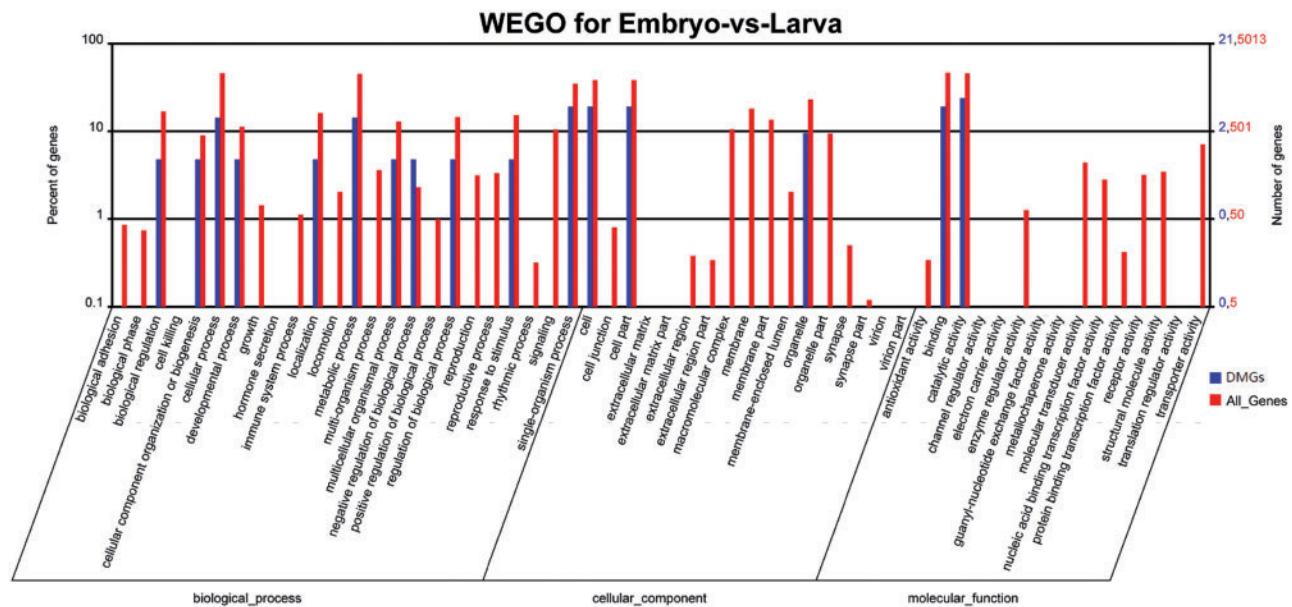


Figure 9. Gene annotation for differentially methylated genes (DMG) between the embryo and larva stages.

methylation level of these sites was >25%, which was much higher than that for the non-CpG methylation sites (Supplementary Table S4). Moreover, most of these methylated sites were enriched in gene regions (Supplementary Tables S5 and S6), as observed for mCG sites in other eukaryotes. Since mCG sites regulate the expression of the corresponding genes in other organisms, we wondered whether this function also occurs in *T. castaneum*. We observed that the average expression levels of these genes exhibiting symmetrical mCpG sites in the gene body were higher than those of the genes lacking mCG sites in the four life stages (Supplementary Fig S18). This finding indicated that the symmetrical mCG sites in *T. castaneum* are positively correlated with gene expression level, similar to the classical functions of mCG in other organisms.

3.10. Validation of DNA methylation in *T. castaneum*

To validate the sequencing results, two different strategies were employed. First, the global methylation levels of genomic DNA in the four life stages were assessed using an immunochemical approach. The four samples showed low but detectable genomic methylation levels. Moreover, the global methylation level exhibited a weakly increasing trend from the embryo to the adult stages, which is consistent with the sequencing results (Supplementary Fig. S19). To check the methylation status at the gene level, methylation-sensitive restriction enzyme-mediated PCR was further conducted. Genomic DNA from the four stages was digested with three methylation-sensitive restriction enzymes: the methylation-sensitive isoschizomers *MspI* and *HpaII*, which recognize the CCGG sequence, and *ClaI*, which recognizes the ATCGAT sequence. Seven methylated regions containing the single CCGG sequence were amplified from the four samples. Genomic DNA digested with *MspI*/*HpaII* was used as a PCR template. The results were consistent with the sequencing results indicating the CpC and CpG methylation patterns of the *T. castaneum* genome (Supplementary Fig. S21A). Four methylated regions with the sole ATCGAT sequence were selected from the four samples. Genomic DNA digested with *ClaI* was used as a PCR template. The cytosine in the ATCGAT sequence was

symmetrically methylated on both strands in the four samples at chromosome 8:8508082(+)/8508083(−) according to the sequencing results; consistent with this result, the amplicons were present in the wells corresponding to *ClaI*-digested DNA in all four samples (Supplementary Fig. S21B). Overall, the two verification experiments illustrate the presence of symmetrical CpG and asymmetrical non-CpG methylation in *T. castaneum*.

4. Discussion

This study provides the first unequivocal evidence of cytosine methylation at specific sites in the four life stages of *T. castaneum* and reveals that two types of cytosine methylation exist in *T. castaneum*: asymmetrical non-CpG methylation and symmetrical CpG methylation.

For non-CpG methylation, we observed four typical features: (1) a predominant distribution of this form of methylation in the *T. castaneum* methylome; (2) a relatively low methylation level of methylated cytosines; (3) accumulation of mCs in intergenic regions and introns but depletion from gene bodies; and (4) heavy methylation of repeat elements.

In eukaryotic organisms, methylated cytosines primarily occur in the symmetrical CpG dinucleotide context.^{3,29,42–44} However, a small quantity of non-CpG methylation (mCHG and mCHH, H stands for the non-G nucleotides) has also been observed in mouse and human embryonic stem cells,⁴⁵ *Camponotus floridanus* and *Harpegnathos saltator*,²⁶ *Trichinella spiralis*,⁴⁰ *Nasonia vitripennis*,²⁸ and *Crassostrea gigas*.²⁷ In the present study, we not only discovered that methylation occurs in various contexts (mCpG, mCpA, mCpC, and mCpT) in *T. castaneum* but also found that non-CpG methylation (mostly mCpA) predominates throughout the methylome, which is consistent with previous results regarding cytosine methylation in *T. castaneum*.³⁵ Moreover, the DNA methylome profiles of *D. melanogaster* and *Polistes canadensis*⁴⁶ show a similar pattern to that of *T. castaneum*, with most of the methylated cytosines being concentrated in non-CpG motifs.³¹ It has been suggested that non-CpG methylation in animals is a by-product of DNMT3, as

this type of methylation has been detected in cells with strong *de novo* methyltransferase activity, such as ESC cells.²⁰ However, there is no DNMT3 gene in the *T. castaneum*, *D. melanogaster*, and *P. canadensis* genomes, which indicates that novel and distinct pathways in these organisms are responsible for non-CpG methylation. Although non-CpG methylation plays important roles in gene silencing, suppression of transposons and genomic imprinting in plants,^{29,47,48} its function in animals remains to be elucidated. Moreover, the small number of mCpG sites present in *T. castaneum* may explain why a previous study did not detect methylated cytosines based on CpG(O/E) analysis.⁴⁵

Among symmetrical mCG sites in the eukaryotic methylome, the methylation level exhibits a typical bimodal distribution. Most of the methylated sites are either highly methylated or unmethylated, which implies that the methylation pattern is uniform among different cell types throughout the body.³ In *T. castaneum*, the methylation level of most methylated cytosines was ~10% (Fig. 3C). A similar pattern has only been observed in the *D. melanogaster* methylome, in which the methylation level of the most highly methylated sites is only 50%, and most of the sites are methylated in <5% of the genome. A relatively low methylation level has also been observed for non-CpG methylated sites in human ESC cells and ants.²⁰ It seems that a low methylation level is a common feature of non-CpG methylation. Combined with the allelic pattern of methylation we identified, the low methylation level found in *T. castaneum* may indicate that methylation is restricted to small subsets of cells or that its genomic distribution is variable from cell to cell. The different distribution patterns between the symmetrical mCpG sites and non-CpG sites may indicate that these two types of DNA methylation are catalysed by distinct enzymatic systems. In addition, the low methylation level observed in *T. castaneum* may explain why the two previous studies using BS-Seq to identify the methylomes of the larva and adult of *T. castaneum* could not identify methylated cytosines: the coverage was too low to identify these cytosines.

According to previous reports, in most eukaryotic organisms, DNA methylation is significantly enriched in gene bodies, especially in transcribed exons.⁴² This appears to be the conserved feature of eukaryotic methylomes. However, the functions of gene body methylation were not previously clear. The proposed functions of gene body methylation include silencing of repetitive DNA elements in gene bodies,⁴⁹ regulation of alternative splicing or alternative promoter silencing,^{11,50} and ensuring accurate transcription.⁵¹ However, in *T. castaneum*, most methylated sites are concentrated in intergenic regions and absent from gene bodies (Fig. 5). A similar pattern has been detected in the *D. melanogaster* methylome,³¹ raising questions about the function of DNA methylation in intergenic regions. Since many functional elements are located in intergenic regions, including enhancers, insulators, and many transcription factor binding sites, methylation in these regions may play important roles in normal eukaryotic development. Interestingly, the average methylation level of repeat elements was found to be higher than that of the genome in *T. castaneum*. Among the different classes of repeat elements, simple-repeat elements exhibited the highest methylation level, similar to what occurs in *D. melanogaster*. In *D. melanogaster*, methylated cytosines primarily occur in CA, CT-rich 4–5-base simple-repeat elements. The different classes of TEs exhibit variable methylation patterns; in certain classes, such as LINEs, the DNA is hypermethylated compared with the average genome level, whereas other TEs, including LTRs and RCs, are hypomethylated. The methylation pattern of transposons varies greatly among different organisms.

TEs are heavily methylated in land plants and vertebrates^{3,29} but are rarely methylated in most insects, such as *Bombyx mori*, *Apis mellifera*, *Nasonia vitripennis*, and *D. melanogaster*^{24,25,28} and are moderately methylated in *Ciona intestinalis*.

Overall, the distribution pattern of non-CpG methylation in *T. castaneum* is novel and unique and does not follow the ancient and conserved features of the eukaryotic methylome but does share high similarity with the *D. melanogaster* methylome. We examined the mechanisms underlying the establishment and maintenance of DNA methylation to obtain insights into the source of this divergence. In eukaryotes, DNA methylation is catalysed by a family of DNMTs, which are divided into three groups: DNMT1, DNMT2, and DNMT3.⁵² Among the DNMTs, DNMT3 initiates *de novo* methylation in the CpG context during gametogenesis and embryogenesis,^{53,54} which is then maintained beyond DNA replication by DNMT1. In contrast to these DNMTs, DNMT2 primarily performs tRNA methylation^{55–57} and plays a lesser role in DNA methylation.⁵⁸ However, the *D. melanogaster* genome only contains the *dnmt2* gene and lacks *dnmt1* and *dnmt3*, which suggests that the traditional methylation system has been lost during the evolution of this species.⁸ A recent study confirmed that DNMT2-deficient *Drosophila* strains do not exhibit decreased DNA methylation levels,³⁰ which indicated that there must be another methyltransferase in the *Drosophila* genome responsible for the novel methylation pattern. This methyltransferase may present no sequence similarity with the traditional methyltransferases or may be located in unsequenced regions of the *Drosophila* genome. *T. castaneum* exhibits two methyltransferase genes, *dnmt1* and *dnmt2*. However, according to our results, *dnmt1* produces two alternative splicing transcripts in *T. castaneum*, with the truncated form, *Tcdnmt1b*, which has lost the catalytic domain of DNA methyltransferases, showing significant abundance among all life stages, in contrast to the conserved form, *Tcdnmt1a*. Although *dnmt1* still exists in *T. castaneum*, its methyltransferase function appears to be significantly weakened. Thus, we suggest that the conserved DNA methylation patterns catalysed by the traditional DNMTs are much weaker or nearly absent in *T. castaneum* and that the detected novel methylation pattern, which shows high similarity to the *Drosophila melanogaster* methylome, may be catalysed by unknown DNMTs, consistent with what occurs in *D. melanogaster*.

However, the presence of *Tcdnmt1a* suggests that the traditional DNA methylation system may still function to methylate genomic cytosines in *T. castaneum*, even though this function is greatly weakened. Therefore, we analysed the symmetrical mCpG sites in the methylomes from four different stages. Although only a small number of symmetrical methylated CpG sites are present in *T. castaneum*, the methylated sites were most prominent within gene bodies and were positively correlated with gene expression, consistent with the conserved features of the eukaryotic methylome.

While the evidence implies that novel methylation does exist in *T. castaneum*, its function is not very clear. The lack of dependence of this type of methylation on conserved DNMTs indicates that the phenotype of methyltransferase mutants cannot be used to infer the functions of DNA methylation. To reveal the importance of DNA methylation, we await the identification of the novel DNMTs. However, we observed a weak, but significant relationship between higher gene body methylation and a lower expression level of the corresponding gene, which suggests that genome methylation in beetles may be involved in gene regulation, as observed in *D. melanogaster*. Moreover, comparison of the methylomes associated with

different developmental stages revealed that the stage-specific DMRs are preferentially located in gene regions and closely associated with differences in gene expression, further supporting the association of DNA methylation and gene expression. Considering the allelic heterogeneity we observed, specific methylation could characterize different subsets of cells throughout the body. If so, methylation in a specific subset of cells may affect expression only in those cells. The influence of methylation on gene expression was not apparent in the present study because the gene expression level was determined from the whole body of *T. castaneum*. The heterocellular pattern is also consistent with the differentiated epigenome. If methylation plays a role in gene regulation, it may exhibit different distribution patterns according to the cell type.⁵⁹

Although the data that we obtained clarify the presence of DNA methylation in *T. castaneum*, it is likely that the methylome is not complete and that we detected only the most highly methylated sites in the genome. There may be other methylated sites in a small subset of the cells that we did not detect. Detecting this type of methylation may require sequencing at extreme depths. Consequently, we identified only a few DMRs between different developmental stages. There may be more DMRs between the different stages that are too rare to be detected. A recent study in *D. melanogaster* suggested that the DNA N6-methyladenine (6mA) modification is present in the *Drosophila* genome and is regulated by the *Drosophila* Tet homolog.⁶⁰ Taken together with the relatively low cytosine methylation level observed in *D. melanogaster*, it is believed that 6mA may fulfil the functions of 5mC. Similarly, considering the low methylation level and the existence of a Tet homologue (TcasGA2_TC013798) in *T. castaneum*³⁵, it is likely that the 6mA modification also exists in *T. castaneum* to fulfil the function of cytosine methylation. Of course, this will wait the future study.

Our study provides a clear answer to the question of whether cytosine DNA methylation is present in *T. castaneum* and reveals that two types of DNA methylation exist in *T. castaneum*: conserved CpG methylation and novel non-CpG methylation. Traditional CpG methylation, in which methylated sites occur symmetrically in the CpG context and are highly methylated, is catalysed by Tcdnmt1a, whereas the novel non-CpG methylation pattern, which shows high similarity to the *D. melanogaster* methylome, may be catalysed by mysterious methyltransferases. Although symmetrical mCG sites are rare throughout the genome, they still regulate gene expression. However, the origin and function of the novel methylation remains unclear. The unknown methyltransferases either share no sequence similarity with known DNMTs or are located in unsequenced regions of eukaryotic genomes. As DNMT2 can perform RNA methylation, it is possible that an RNA methyltransferase may act on DNA, although there is no evidence that such activity occurs. The heterocellular distribution of non-CpG methylation may imply that the methylated sites play roles only in the corresponding cells. Additional studies to detect the methylomes of specific tissues or cells are needed to resolve this issue. The pattern of DNA methylation observed in *T. castaneum* and *D. melanogaster* represents a new world of DNA methylation in eukaryotes and indicates that the variety of DNA methylation in eukaryotes is much greater than previously thought. New technologies to detect the methylomes of more organisms with much deeper coverage or to confirm the novel methyltransferase responsible for this unique form of DNA methylation will reveal more functions of DNA methylation.

Availability of data and materials

The Tcdnmt1a, Tcdnmt1b, and Tcdnmt2 mRNA sequences have been submitted to the GenBank. The corresponding accession numbers are as follows: Tcdnmt1a accession = KX553973, Tcdnmt1b accession = KX553974, and Tcdnmt2 accession = KX553975. Sequence data and processed data from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE 84253. This comprises the subseries GSM2230370 Embryo (BS-Seq), GSM2230371 Larva (BS-Seq), GSM2230372 Pupa (BS-Seq), GSM2230373 Adult (BS-Seq), GSM2230374 Embryo (RNA-Seq), GSM2230375 Larva (RNA-Seq), GSM2230376 Pupa (RNA-Seq), and GSM2230377 Adult (RNA-Seq).

Supplementary data

Supplementary data are available at DNARES online.

Funding

This work was supported by the National Natural Science Foundation of China (Nos. 31572326, 31172146, and 31601893), the PAPD of Jiangsu Higher Education Institutions and General Project of Natural Science Foundation of the Jiangsu high Education Institutions (No. 16KJB180004).

Conflict of interest

None declared.

References

1. He, X. J., Chen, T. P. and Zhu, J. K. 2011, Regulation and function of DNA methylation in plants and animals, *Cell Res.*, **21**, 442–65.
2. Suzuki, M. M. and Bird, A. 2008, DNA methylation landscapes: provocative insights from epigenomics, *Nat. Rev. Genet.*, **9**, 465–76.
3. Zemach, A., McDaniel, I. E., Silva, P. and Zilberman, D. 2010, Genome-wide evolutionary analysis of eukaryotic DNA methylation, *Science*, **328**, 916–9.
4. Klose, R. J. and Bird, A. P. 2006, Genomic DNA methylation: the mark and its mediators, *Trends Biochem. Sci.*, **31**, 89–97.
5. Heard, E., Clerc, P. and Avner, P. 1997, X-chromosome inactivation in mammals, *Ann. Rev. Genet.*, **31**, 571–610.
6. Li, E., Beard, C. and Jaenisch, R. 1993, Role for DNA methylation in genomic imprinting, *Nature*, **366**, 362–5.
7. Reik, W. 2007, Stability and flexibility of epigenetic gene regulation in mammalian development, *Nature*, **447**, 425–32.
8. Goll, M. G. and Bestor, T. H. 2005, Eukaryotic cytosine methyltransferases, *Annu. Rev. Biochem.*, **74**, 481–514.
9. Bird, A. 2002, DNA methylation patterns and epigenetic memory, *Genes Dev.*, **16**, 6–21.
10. Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T. and Henikoff, S. 2007, Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription, *Nat. Genet.*, **39**, 61–9.
11. Shukla, S., Kavak, E., Gregory, M., et al. 2011, CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing, *Nature*, **479**, 74–9.
12. Walsh, C. P., Chaillet, J. R. and Bestor, T. H. 1998, Transcription of IAP endogenous retroviruses is constrained by cytosine methylation, *Nat. Genet.*, **20**, 116–7.
13. Zhang, X., Yazaki, J., Sundaresan, A., et al. 2006, Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*, *Cell*, **126**, 1189–201.

14. Rhee, I., Bachman, K. E., Park, B. H., et al. 2002, DNMT1 and DNMT3b cooperate to silence genes in human cancer cells, *Nature*, **416**, 552–6.
15. Szyf, M. 2008, The role of DNA hypermethylation and demethylation in cancer and cancer therapy, *Curr. Oncol.*, **15**, 72.
16. Szyf, M., Pakneshan, P. and Rabbani, S. A. 2004, DNA methylation and breast cancer, *Biochem. Pharmacol.*, **68**, 1187–97.
17. De Smet, C. and Lioriot, A. 2010, DNA hypomethylation in cancer: epigenetic scars of a neoplastic journey, *Epigenetics*, **5**, 206–13.
18. Ting, A. H., McGarvey, K. M. and Baylin, S. B. 2006, The cancer epigenome—components and functional correlates, *Genes Dev.*, **20**, 3215–31.
19. Su, Z., Han, L. and Zhao, Z. 2011, Conservation and divergence of DNA methylation in eukaryotes: new insights from single base-resolution DNA methylomes, *Epigenetics*, **6**, 134–40.
20. Laurent, L., Wong, E., Li, G., et al. 2010, Dynamic changes in the human methylome during differentiation, *Genome Res.*, **20**, 320–31.
21. Deaton, A. M. and Bird, A. 2011, CpG islands and the regulation of transcription, *Genes Dev.*, **25**, 1010–22.
22. Jones, P. A. and Takai, D. 2001, The role of DNA methylation in mammalian epigenetics, *Science*, **293**, 1068–70.
23. Lister, R., Pelizzola, M., Dowen, R. H., et al. 2009, Human DNA methylomes at base resolution show widespread epigenomic differences, *Nature*, **462**, 315–22.
24. Xiang, H., Zhu, J., Chen, Q., et al. 2010, Single base-resolution methylome of the silkworm reveals a sparse epigenomic map, *Nat. Biotechnol.*, **28**, 516–20.
25. Lyko, F., Foret, S., Kucharski, R., Wolf, S., Falckenhayn, C. and Maleszka, R. 2010, The honey bee epigenomes: differential methylation of brain DNA in queens and workers, *PLoS Biol.*, **8**, 2834.
26. Bonasio, R., Li, Q. Y., Lian, J. M., et al. 2012, Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and *Harpegnathos saltator*, *Curr. Biol.*, **22**, 1755–64.
27. Wang, X. T., Li, Q. Y., Lian, J. M., et al. 2014, Genome-wide and single-base resolution DNA methylomes of the Pacific oyster *Crassostrea gigas* provide insight into the evolution of invertebrate CpG methylation, *BMC Genomics*, **15**, 1119.
28. Wang, X., Wheeler, D., Avery, A., et al. 2013, Function and evolution of DNA methylation in *Nasonia vitripennis*, *PLoS Genet.*, **9**, e1003872.
29. Feng, S., Cokus, S. J., Zhang, X., et al. 2010, Conservation and divergence of methylation patterning in plants and animals, *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 8689–94.
30. Raddatz, G., Guzzardo, P. M., Olova, N., et al. 2013, Dnmt2-dependent methylomes lack defined DNA methylation patterns, *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 8627–31.
31. Takayama, S., Dhahbi, J., Roberts, A., et al. 2014, Genome methylation in *D. melanogaster* is found at specific short motifs and is independent of DNMT2 activity, *Genome Res.*, **24**, 821–30.
32. Law, J. A. and Jacobsen, S. E. 2010, Establishing, maintaining and modifying DNA methylation patterns in plants and animals, *Nat. Rev. Genet.*, **11**, 204–20.
33. Glastad, K. M., Hunt, B. G., Yi, S. V. and Goodisman, M. A. D. 2011, DNA methylation in insects: on the brink of the epigenomic era, *Insect Mol. Biol.*, **20**, 553–65.
34. Cunningham, C. B., Ji, L., Wiberg, R. A., et al. 2015, The genome and methylome of a beetle with complex social behavior, *Nicrophorus vespilloides* (Coleoptera: Silphidae), *Genome Biol. Evol.*, **7**, 3383–96.
35. Feliciello, I., Parazajder, J., Akrap, I. and Ugarkovic, D. 2013, First evidence of DNA methylation in insect *Tribolium castaneum* environmental regulation of DNA methylation within heterochromatin, *Epigenetics*, **8**, 534–41.
36. Beeman, R. W., Stuart, J. J., Brown, S. J. and Denell, R. E. 1993, Structure and function of the homeotic gene complex (HOM-C) in the beetle, *Tribolium castaneum*, *BioEssays*, **15**, 439–44.
37. Li, C., Zhang, Y., Yun, X., et al. 2014, Methuselah-like genes affect development, stress resistance, lifespan and reproduction in *Tribolium castaneum*, *Insect Mol. Biol.*, **23**, 587–97.
38. Kim, H. S., Murphy, T., Xia, J., et al. 2010, BeetleBase in 2010: revisions to provide comprehensive genomic information for *Tribolium castaneum*, *Nucleic Acids Res.*, **38**, D437–42.
39. Richards, S., Gibbs, R. A., Weinstock, G. M., et al. 2008, The genome of the model beetle and pest *Tribolium castaneum*, *Nature*, **452**, 949–55.
40. Gao, F., Liu, X. L., Wu, X. P., et al. 2012, Differential DNA methylation in discrete developmental stages of the parasitic nematode *Trichinella spiralis*, *Genome Biol.*, **13**, R100.
41. Suzuki, M. M., Kerr, A. R. W., De Sousa, D. and Bird, A. 2007, CpG methylation is targeted to transcription units in an invertebrate genome, *Genome Res.*, **17**, 625–31.
42. Lee, T.-F., Zhai, J. and Meyers, B. C. 2010, Conservation and divergence in eukaryotic DNA methylation, *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 9027–8.
43. Golbabapour, S., Abdulla, M. A. and Hajrezaei, M. 2011, A concise review on epigenetic regulation: insight into molecular mechanisms, *Int. J. Mol. Sci.*, **12**, 8661–94.
44. Henderson, I. R. and Jacobsen, S. E. 2007, Epigenetic inheritance in plants, *Nature*, **447**, 418–24.
45. Ramsahoye, B. H., Biniszkiwicz, D., Lyko, F., Clark, V., Bird, A. P. and Jaenisch, R. 2000, Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a, *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 5237–42.
46. Patalano, S., Vlasova, A., Wyatt, C., et al. 2015, Molecular signatures of plastic phenotypes in two eusocial insect species with simple societies, *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 13970–5.
47. Stroud, H., Do, T., Du, J., et al. 2014, Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis, *Nat. Struct. Mol. Biol.*, **21**, 64–72.
48. Lindroth, A. M., Cao, X., Jackson, J. P., et al. 2001, Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation, *Science*, **292**, 2077–80.
49. Yoder, J. A., Walsh, C. P. and Bestor, T. H. 1997, Cytosine methylation and the ecology of intragenomic parasites, *Trends Genet.*, **13**, 335–40.
50. Maunakea, A. K., Nagarajan, R. P., Bilenky, M., et al. 2010, Conserved role of intragenic DNA methylation in regulating alternative promoters, *Nature*, **466**, 253–7.
51. Smith, Z. D. and Meissner, A. 2013, DNA methylation: roles in mammalian development, *Nat. Rev. Genet.*, **14**, 204–20.
52. Bestor, T. H. 2000, The DNA methyltransferases of mammals, *Hum. Mol. Genet.*, **9**, 2395–402.
53. Okano, M., Xie, S. and Li, E. 1998, Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases, *Nat. Genet.*, **19**, 219–20.
54. Kafri, T., Ariel, M., Brandeis, M., et al. 1992, Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line, *Genes Dev.*, **6**, 705–14.
55. Goll, M. G., Kirpekar, F., Maggert, K. A., et al. 2006, Methylation of tRNA^{Asp} by the DNA methyltransferase homolog Dnmt2, *Science*, **311**, 395–8.
56. Schaefer, M., Pollex, T., Hanna, K., et al. 2010, RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage, *Genes Dev.*, **24**, 1590–5.
57. Jurkowski, T. P., Meusburger, M., Phalke, S., et al. 2008, Human DNMT2 methylates tRNA^{Asp} molecules using a DNA methyltransferase-like catalytic mechanism, *RNA*, **14**, 1663–70.
58. Hermann, A., Schmitt, S. and Jeltsch, A. 2003, The human Dnmt2 has residual DNA-(cytosine-C5) methyltransferase activity, *J. Biol. Chem.*, **278**, 31717–21.
59. Boffelli, D., Takayama, S. and Martin, D. I. 2014, Now you see it: genome methylation makes a comeback in Drosophila, *BioEssays*, **36**, 1138–44.
60. Zhang, G., Huang, H., Liu, D., et al. 2015, N6-methyladenine DNA modification in Drosophila, *Cell*, **161**, 893–906.