# Genome-wide DNA Methylation Patterns in *Daphnia magna* Reveal Targeting to Promoter Proximal Regions


## Authors

Ruoshui Liu[1], Marco Morselli[2], Lev Y. Yampolsky[3], Leonid Peshkin*[4], Matteo Pellegrini*[1]

[1] Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, CA 90024, USA

[2] Department of Chemistry, Life Sciences, and Environmental Sustainability, University of Parma, Parma, Italy

[3] Department of Biological Sciences, East Tennessee State University, Johnson City, TN 37614, USA

[4] Systems Biology, Harvard Medical School, Boston, MA 02115, USA

* Corresponding Authors

pesha@hms.harvard.edu

matteop@ucla.edu



## Abstract

Studying DNA methylation in *Daphnia magna*, a model organism in ecological and evolutionary research, offers valuable insights into pharmaceutical toxicity and behavioral ethology. In this study, we characterized DNA methyltransferases and mapped DNA methylation across the *D. magna* genome using whole-genome bisulfite sequencing. Our analysis revealed a highly expressed, nonfunctional *de novo* methyltransferase (DNMT3.1) alongside lowly expressed functional DNMT3.2 and maintenance methyltransferase DNMT1. *D. magna* displays overall low DNA methylation, targeting primarily CpG dinucleotides. Methylation is sparse at promoters but elevated in the first exons downstream of transcription start sites, with these exons showing hypermethylation relative to adjacent introns. In contrast to prior studies, we observed minimal age-related changes in DNA methylation patterns that were not sufficiently robust to build an accurate epigenetic clock. These findings expand our understanding of the epigenetic landscape in *D. magna*.



## Keywords

*Daphnia magna*; DNA methylation; epigenetic clock; DNA methyltransferase (DNMT)

## INTRODUCTION

*Daphnia* are planktonic crustaceans within the Phyllopoda subclass and the Branchiopoda class (1). Among over 100 species within the *Daphnia* genus, *Daphnia magna* (*D. magna*) is widely recognized as a model organism in biological research, including studies on pharmaceutical toxicity, reproductive cycles, behavioral ethology, and phenotypic plasticity (2). DNA methylation, the addition of a methyl group on the 5' carbon of cytosine, plays a role in transcriptional regulation and phenotypic expression (3). Previous studies have demonstrated the effect of compound exposure on DNA methylation (4). Given the significance of *D. magna* as a model organism, examining its epigenomics landscape, particularly DNA methylation pattern and regulation, could be critical (5).

CpG methylation is closely associated with gene silencing and is crucial for development (6,7). In *Daphnia magna*, cytosine methylation levels are relatively low, averaging around 1% across the genome, yet specific regions, such as exons, exhibit higher methylation, with exons 2 to 4 displaying the highest levels (3,6,8). Previous studies, however, have tended to average methylation across different genes for each exon, overlooking the nuances of individual methylation. To address this gap, our study aims to examine methylation patterns more granularly and investigate the exon-level and gene-level methylation variations across the genes.

DNA methyltransferases (DNMTs) are a family of enzymes that epigenetically regulate gene expression by establishing and maintaining CpG methylation patterns (9). Maintenance methyltransferase DNMT1 replicates DNA methylation patterns to daughter cells during cell divisions, while the functional *de novo* methyltransferase DNMT 3 establishes methyl groups at unmethylated cytosine sites (10,11). Our research aims to explore the functional domains of *D. magna*'s DNMTs and their effect on genome-wide and gene-specific methylation patterns.

DNA methylation levels vary with age and influence the functional capability of organs in mammals, positioning DNA methylation-based biomarkers, also known as epigenetic clocks, as effective estimators of biological age (12–14). Epigenetic clocks facilitate understanding the impact of both endogenous (epigenetic drift, etc.) and exogenous stressors on biological aging by comparing epigenetic age to chronological age (12,14). In a previous study, an epigenetic clock for *D. magna* was built with 12 clock CpGs (6). However, this analysis was conducted using samples from only two age groups: 10-day and 50-day-old. Additionally, they incorporated *Daphnia* specimens from two different strains and included an interaction term for strain and age in their epigenetic clock. Both the large age gaps and an additional factor of strain potentially diminished the predictive accuracy of their regression model (6).

Thus, further research is needed to characterize age-related changes in DNA methylation and determine the validity of an epigenetic clock for *D. magna*.

This study aimed to investigate the targeting and age association of DNA methylation in *D. magna*. We cultured 17 *D. magna* samples of various ages, performed whole genome bisulfite sequencing, and characterized the DNA methylation patterns at the genome-wide, gene, and exon levels. To better understand the methylation patterns, we analyzed functional domains and mRNA expression levels of DNA methyltransferases. Lastly, we examined age-related changes in methylation patterns and attempted to build an epigenetic clock.

## MATERIALS AND METHODS

### Sample collection

A heat-tolerant clone of D. magna, IL-MI-8, was originally acquired from the Ebert Lab at the University of Basel, Switzerland. This clone was sourced from a pond in Jerusalem, Israel, as part of the Ebert Lab stock collection. To ensure synchronized cohorts, neonates born within a two-day window were separated from their mothers. Sex was determined at 8- to 9-day-old animals. Female individuals were exclusively used in all experiments conducted in this study. All mothers received adequate nutrition and were cultured under identical conditions at 25°C within an incubator. All cultures, including mother cultures, neonate cultures, and tank cultures, were maintained in ADaM water (15). Animals were exposed to a light cycle consisting of 16 hours of light followed by 8 hours of darkness. Animals were fed a daily suspension of the green alga Scenedesmus obliquus at a concentration of 10^5 cells/ml (adjusted for population density, with one animal per 20 ml). Every sixth day, the water was replaced, and offspring were manually removed until animals were transferred to the culture platform. The operational protocols for the culture platform are akin to those described in previously published protocols (16).

### DNA extraction and WGBS library preparation

Approximately 30-50 ng of purified genomic DNA in 50 µl has been subject to sonication using the Bioruptor Pico (Diagenode) for 15 cycles (30 ON; 90 OFF). Fragmentation patterns have been checked with the D1000HS Assay (Agilent Technologies) on a 4200 TapeStation. The volume of the fragmented DNA has been reduced to 20 µl using a Vacufuge (Eppendorf), then subject to bisulfite conversion using the EZ DNA Methylation-Lightning Kit (Zymo Research). The libraries have been prepared using the Accel-NGS Methyl kit (Swift Biosciences - now xGen Methyl-Seq Library Prep - IDT) according to the manufacturer's recommendations except for a major modification. Briefly, the denatured BS-converted gDNA was subject to the adaptase, extension, and ligation reaction. Following the ligation purification, the DNA underwent primer extension (98C for 1 minute; 70C for 2 minutes; 65C for 5 minutes; 4C hold) using oligos containing random UMI at the location of the i5 barcode. The extension using a UMI-containing primer allows the tagging of each molecule to remove PCR duplicates and correctly estimate DNA methylation levels. Following exonuclease I treatment and subsequent purification, the libraries were then amplified using a universal custom P5 primer and i7-barcoded P7 primers (initial denaturation: 98C for 30 seconds; 10 cycles of 98C for 10 seconds, 60C for 30 seconds, 68C for 60 seconds; final extension: 68C for 5 minutes; 4C hold). The resulting single-indexed libraries were then purified and quantified using the Qubit HS-dsDNA assay, and the quality was checked using

the D1000-HS assay (Agilent - TapeStation 4200). The libraries were sequenced as 150 PE on the Illumina NovaSeq6000 platform.

**Bisulfite sequencing data processing**

An alignment index incorporating a bisulfite-converted sequence for each reference strand was constructed from the *Daphnia magna* Xinb3 reference genome (BioProject ID: PRJNA624896, D. Ebert, personal communication) and the mitochondrial genome (17), using the *BSBolt Index* tool (18). The FASTQ files were aligned using *BSBolt Align* (default options) (18). PCR duplicates were removed with *samtools markdup* v1.17 (option -r) (19). Sequence alignment quality was assessed using the CIGAR (compact idiosyncratic gapped alignment report) strings, and sequences with a total number of matches below 50 were excluded from further analysis using *samtools view* and *samtools index* (19). The number of mapped reads in each sample was quantified using *samtools flagstat* (19).

DNA methylation calling was performed with *BSBolt CallMethylation* v1.6.0 (options: -IO) with the DNA alignment index as a reference, generating CGmap files (18,20). The CGmap files were subsequently filtered based on contig coverage and quality. Average coverage before and after filtering for cytosines across each chromosome was obtained using *CGmaptools mec* (20). Global DNA methylation levels were analyzed with *CGmaptools mstat* (20), and differences in global methylation levels of CpG and non-CpG cytosines were assessed using the Wilcoxon test. DNA methylation levels of genes, exons, and introns were calculated using *CGmaptools mtr* (20). Metagene plots for genes and exons were generated using *CGmaptools bed2fragreg* and *CGmaptools mfg* (20). Pathway enrichment analysis was performed using the R package *clusterProfiler* (21). Genome browser snapshots were taken from Integrative Genomics Viewer (version IGV_2.18.2) (22).

The matrix of common CpG sites was generated using *BSBolt AggregateMatrix* function (options: -min-coverage 10 -min-sample 0.8 -CG), with parameters set to a minimum coverage of 10 and an inclusion threshold of at least 80% across the 17 samples (18). To handle missing data, CpG sites absent in some samples were imputed using the *BSBolt Impute* feature. UMAP was employed to visualize the clustering of samples based on methylated CpG sites. This approach allowed us to examine whether *Daphnia* samples from the same age group exhibited consistent methylation patterns, with an expectation of tighter clustering within age groups in the UMAP plot (23).

The CpG sites were filtered to only include the ones with the top 20% of the variability. The epigenetic clock model was built with the Lasso function from the Python module *sklearn.linear_model* and Leave-One-Out Cross Validation (24). Random permutation was used to assess the statistical

significance of differences between age groups at each meta-gene ridge plot position. Observed mean differences between groups were compared to a null distribution generated by 100,000 random permutations. P-values were calculated as the proportion of permuted differences as extreme as the observed, using a two-tailed test. Differentially methylated regions (DMR) analysis was performed using the *Metilene* package (options: -m 10 -d 0.001 -t 4 -f 1) (25).

**RNA sequencing data processing**

Details of an age-specific RNAseq experiment are reported elsewhere (26). Briefly, RNA was extracted from somatic tissues (whole bodies) of *Daphnia magna* females from laboratory clone GB-EL75-69 (Basel University *Daphnia* Stock Collection, Switzerland) of different ages: young, reproducing (age 15-20 days), old, reproductively senescent, and old, reproductively rejuvenated (age both 130-175 days), in 4 replicates each. In all cases**,** *Daphnia* samples were sampled within 24 hours of molting and egg-laying and freshly laid eggs (if any) were removed from the brood chamber before homogenizing *Daphnia*. RNAs were extracted using Qiagen RNeasyPlus Mini kit and RNA library preparation was performed using NEBNext Ultra II DirecMonal RNA Library Prep Kit (NEB, Lynn, MA) following the manufacturer's protocols. The libraries were sequenced with Illumina Novoseq 6000, S4 flow cell, PE100. Reads were mapped to D. magna Xinb3 reference transcriptome (BioProject ID: PRJNA624896; D. Ebert and P. Fields, personal communication), and genes with differential expression either between young vs. old *Daphnia* were identified using DEseq2 (27). Statistical analysis of RNA expression levels between groups was performed using a two-sided t-test.

**DNA methyltransferase analysis**

The functional analysis of DNMT3.1, DNMT3.2, DNMT1, and UHRF1 was conducted using InterProScan (28). Gene sequences were input into the tool to identify and characterize the functional domains within these methyltransferases.

## RESULTS

### *D. magna* has three DNA methyltransferases

Based on the homology with other organisms' DNA methyltransferases (DNMTs), *D. magna* harbors two functional and one non-functional DNA methyltransferases (**Table S1**) (9). It has two *de novo* DNMTs, *Dma*DNMT3.1 and *Dma*DNMT 3.2 (**Figure 1A**) (10).

To better understand DmaDNMT and its relationship with the species' methylation pattern, we used mammalian (human) DNMTs as a reference and compared the homologies between *Dma*DNMTs and *Hs*DNMTs. Mammalian DNMT3a/b contains a PWWP domain, an ADD domain, and a methyltransferase (MTase) domain with 6 motifs (I to VI). Non-functional *Dma*DNMT3.1 is missing 5 critical motifs in the MTase domain, and it is thought to exert similar functions to mammalian DNMT3L. Both *Dma*DNMT3.1 and *Hs*DNMT3L contain putatively functional ADD (ATRX-DNMT3A-DNMT3L) domains that do not bind H3K4me3. However, only *Dma*DNMT3.1 owns the PWWP domain (Pro-Trp-Trp-Pro) that is known to bind methylated lysines, specifically H3K36me2/3 in mammals (11).

Conversely, the "active" *de novo* DNA methyltransferase, *Dma*DNMT3.2, contains all six DNA methyltransferase motifs and the ADD domain, suggesting that it can sense the methylation status of H3K4 and bind unmethylated H3K4 histones (10). However, it is missing the N-terminal PWWP domain present in the active mammalian DNMT3A and DNMT3B.

The other putatively active DNA methyltransferase in *D. magna* is the maintenance methyltransferase *Dma*DNMT1. Compared to the mammal counterpart, such as *Hs*DNMT1, it is missing the N-terminal domains involved in the targeting of the replication foci (DMAP1-BD and PCNA-BD) and a clear nuclear localization signal (NLS). A phosphatase-specific domain (PAP2-wunen) at the N-terminal region is unique to *D. magna* DNMT1.

To understand the activity of the three *D.magna* DNMTs, we analyzed RNA sequencing data for DmaDNMTs across two age groups: young and old (**Figure 1B**). The gene expression levels (TPM) for the three DmaDNMTs varied significantly by a two-sided t-test. The two functional enzymes, *Dma*DNMT1 and *Dma*DNMT3.2, exhibited consistently low expression levels with no significant differences between the groups. By contrast, the non-functional *Dma*DNMT3.1 showed significantly higher expression in the young group compared to the old group. However, since DmaDNMT3.1 lacks the critical DNA methyltransferase motifs, its expression is unlikely to impact methylation levels.

**High-coverage *D. magna* Methylomes**

To determine the DNA methylation profiles of *Daphnia magna*, we obtained two 45-day-old (mature-old) Xinb3 clone individuals, Dap_S1 and Dap_S2. Whole-genome bisulfite sequencing was performed as described in Materials and Methods, and subsequent sequence alignment to the *D. magna* genome and methylation calling was carried out. On average, approximately 79% of the reads mapped primarily to the genome, and this dropped to 56% after removing duplicates and 53% after filtering by CIGAR (compact idiosyncratic gapped alignment report) strings. Applying a minimum coverage threshold of 10 reads per site in the methylation calling step, we identified roughly 41.1 million cytosine dinucleotides across the *D. magna* genome. Although the average coverage across contigs was around 60, the distribution demonstrated large variations (**Figure S1A**).

The original genome assembly of *Daphnia magna* consisted of one mitochondrial genome (17) and 608 nuclear contigs (BioProject ID: PRJNA624896, D. Ebert, personal communication), many of which are short. Approximately 50% of these contigs are under 20,000 base pairs and 25% are below 10,000 base pairs in length. Some short contigs showed pronounced high coverage and hypermethylation at CpG sites, suggesting that they are outliers and possible contaminants. To improve the quality of methylome, alignment filtration was implemented, retaining only the CpN sites from the top-performing 97 nuclear contigs validated by genetic recombination map, Hi-C data, and optical mapping (BioProject ID: PRJNA624896, D. Ebert, personal communication). After selective filtration, 38.7 million methylcytosine sites remained. The mean contig length was 1.3 million base pairs, with an average coverage maintained approximately 60x (**Figure S1B**). This process reduced variability across contigs and confirmed the reliability of the remaining methylcytosine sites.

Besides the two deeply sequenced *D. magna* samples (Dap_S1 and Dap_S2), an additional 15 *D. magna* samples denoted Dap_D1 to Dap_D16 (Dap_D7 excluded due to low quality) were bisulfite sequenced across different age groups: young (9 days), mature (22 to 27 days), and old (51 to 58 days). Similar approaches of alignment, methylation calling, and filtration were applied to the 15 samples (**Table S2**).

**Mitochondrial genome methylation as a negative control for non-conversion artifacts**

Bisulfite treatment converts unmethylated cytosines to uracils, while leaving the methylated cytosines unchanged, allowing methylation to be accurately detected through sequencing. However, bisulfite conversion is not 100% effective, and a small fraction (typically around 1%) of unmethylated cytosines fail to deaminate and are mistakenly identified as methylated(29).

To distinguish between methylcytosines and unconverted cytosines we considered several criteria. If cytosine dinucleotides are methylated, we expect the methylation patterns at individual sites to be consistent across samples. To investigate this, we conducted correlation analyses between the Dap_S1 and Dap_S2 samples across different cytosine contexts. We observed a high correlation (r = 0.949) for CpG sites between the two samples, indicating reproducible methylation patterns. We also expect that methylated CpG dinucleotides are present on both strands, which is often referred to as symmetric methylation. In Dap_S1, we identified 1,100,329 pairs of CpG sites with non-zero methylation on both strands, of which 1,040,273 pairs had a difference in methylation of less than 0.05. Similarly, in Dap_S2, 1,204,966 CpG pairs had non-zero methylation, with 1,129,724 pairs showing a difference below 0.05. This high degree of symmetry between strands further supports the reliability of CpG methylation signals (30). By contrast, non-CpG contexts exhibited much lower correlations, with CpA at r = 0.338, CpC at r = 0.222, and CpT at r = 0.249. The low correlation levels for non-CpG sites suggest that these are likely artifacts of incomplete bisulfite conversion rather than true methylation.

Moreover, to differentiate between bisulfite non-conversion and true cytosine methylation, we used the mitochondrial genome as a negative control, as previous studies have found that mitochondrial DNA is generally unmethylated (31). Cytosine dinucleotides in the mitochondrial genome exhibited median coverage levels exceeding 100x (**Figure S2**) and this high coverage allows us to use the mitochondrial genome as a negative control. Dap_S1 and Dap_S2 displayed consistent methylation patterns across all cytosine contexts (**Figure S3**) with median methylation levels ranging from 0.2% to 0.5%. Non-CpG sites had third-quartile methylation levels between 1.0% and 1.2%, while CpG sites exhibited a slightly higher third quartile of 1.6%. To estimate the non-conversion rate in the nuclear genome, we compared the methylation levels of individual nuclear cytosine sites with those in the mitochondrial genome. The average methylation levels in the mitochondrial genome were comparable to or even higher than those observed in the nuclear genome (**Figure S4**). Moreover, non-CpG sites in the mitochondrial genome showed a higher percentage of methylation levels exceeding 20% compared to nuclear sites. However, the presence of nuclear mitochondrial sequences (nuMT) (32) in *D. magna*—mitochondrial DNA fragments inserted into the nuclear genome—may explain some of these effects (33). These nuMT sequences are fragments of mitochondrial DNA and thus may lead to false positive methylation patterns in the mitochondria. In contrast to non-CpG sites, CpG sites in the nuclear genome displayed higher methylation levels than those in the mitochondrial genome, suggesting that CpG sites in the nuclear genome are methylated rather than artifacts of non-conversion.

### *D. magna* DNA CpG sites are methylated at low levels

To elucidate methylation at a global level, we computed the average methylation levels of the 97 contigs. CpT and CpC sites exhibited an average methylation level of approximately 1.25%, while CpA methylation was slightly lower with 1.1% methylation (**Figure S5**). By contrast, CpG sites displayed higher average methylation levels, around 2.05% across contigs, whose distribution of methylation was significantly different from that of the non-CpG sites (p-value < 0.0001 in Wilcox test).

We further assessed the methylation levels at individual CpN sites and categorized corresponding fractions of methylation levels (**Figure 2**). We found that the two samples had comparable methylation levels, and the methylation levels of CpA, CpT, and CpC sites were distributed similarly. Around 40% of the non-CpG sites had a methylation level below 0.05, while around 55% had no methylation. Approximately 5% of the non-CpG sites had a methylation level above 0.05 and less than 0.2% above 0.20. By contrast, CpG sites are significantly hypermethylated compared to the three non-CpG cytosine contexts in both samples (p-value < 0.0001 in the Wilcox test). Approximately 6% of the CpG sites had a methylation level above 0.05 and around 1% of the sites above 0.20. Specifically, 0.5% of the CpG sites have a methylation level above 0.80. This higher methylation frequency in CpG sites could be attributed to the specificity of DNMTs, which primarily target CpG dinucleotides over non-CpG sites (34). This pattern emphasizes the established role of CpG methylation in invertebrate transcriptional regulation (35), further distinguishing it from non-CpG methylation contexts.

In addition to examining the overall methylation trends, we analyzed the methylation profiles within specific groups of sequences, including DNA transposons, retrotransposons, LTR retrotransposons, and non-LTR elements. While some short interspersed nuclear elements (e.g., SINE_U, SINE_B2, SINE_Alu) showed localized methylation upstream or downstream of the sequences, most repetitive elements exhibited constant low methylation levels.

### *D. magna* DNA CpG sites are hypomethylated upstream of TSS

We next examined methylation patterns around genes, focusing on the range of 3 kilobases upstream of the transcription start site (TSS) to 3 kilobases downstream of the transcription end site (TES). Both samples, Dap_S1 and Dap_S2, showed consistent trends, highlighting the reproducibility of these patterns. Non-CpG sites maintained a uniform methylation level of approximately 1.5% across the gene and proximal regions (**Figure 3A, Figure 3B, Figure 3D**), suggesting a minimal role in gene regulation. In contrast, CpG methylation displayed greater variability at different positions along genes. Specifically, CpG methylation levels were lowest a few hundred bases upstream of the TSS (1.5%

methylated) and sharply increased downstream of the TSS, peaking at nearly 4.5% (**Figure 3C**). Subsequently, within the gene, CpG methylation levels gradually decreased, reaching 1.5%, underscoring the enrichment on the 5' end of the genes.

To minimize positional artifacts from normalizing genes to a uniform distance and to provide more insight into dynamic CpG methylation within genes, we grouped genes into deciles based on mean methylation levels (e.g., top 10%, 10-20%) and calculated methylation levels centered at the TSS and TES. Dap_S1 (**Figure 4**) and Dap_S2 (**Figure S6**) displayed high methylation in the top 20% of genes (**Figure S7**) and minimal methylation in the lower 50% of genes (**Figure S8**). Around the TSS, the top 20% of genes showed markedly higher methylation levels than the rest, with methylation at the 5' end exceeding 30% in the top 10% and nearly reaching 20% in the 10-20% group. At the same time, methylation in the top 10% of genes peaked at 10% upstream of the TES.

We explored the functions of the highly methylated genes and unmethylated genes using pathway enrichment analysis. Genes within the top 10% of methylation in both samples were significantly enriched for fundamental cellular processes in the Gene Ontology Biology Process, including ribonucleoprotein complex biogenesis, cytoplasmic translation, mitochondrial gene expression, respiratory electron transport chain, methylation, and DNA replication. Unmethylated genes showed no enrichment for any specific pathways.

### *D. magna* Exon CpG sites are more methylated than neighboring intron regions

To further explore the landscape of genomic methylation, we analyzed differential methylation patterns between exons and introns. Our findings reveal that exons contain a higher density of methylated CpG sites (**Figure S9**). On average, exons showed a methylation rate of approximately 0.17%, whereas nearby intronic regions were significantly less methylated, averaging 0.07% ($p < 0.001$ in the Wilcox test). Among the exons with a mean methylation rate greater than zero, the average methylation rate is 8.7%, significantly higher than the 4.9% observed in methylated introns ($p < 0.001$ in Wilcoxon test).

We also analyzed methylation levels across exons in genes with at least five exons to understand how exon position affects methylation. We observed that methylation varied significantly with exon positions (**Figure 5A**). Specifically, Exon 1 shows a gradual increase in CpG methylation, peaking near its 3' end, followed by a sharp drop at the transition to the adjacent intron. This pattern of exon hypermethylation and intron hypomethylation is primarily driven by the top 10% of methylated genes (**Figure 5B**). Methylation levels continued to rise from Exon 2 through Exon 4, reaching a peak of nearly 8% in Exon 4 before decreasing (**Figure 5A**).

***D. magna*'s global and local CpG methylation patterns are weakly associated with age**

To determine how the methylation pattern changes with age, we first conducted a global analysis of CpG methylation levels across three age groups: young, mature, and old. The average methylation levels were 3.426% in the young group, 3.386% in the mature group, and 3.311% in the old group, showing a slight decrease with age; however, none of these differences reached statistical significance (**Figure 6A**). We also aggregated CpG methylation values for sites present in at least 80% of samples, imputed missing values, and constructed a CpG matrix. Visualization of this matrix using UMAP did not reveal clear clustering by age group, indicating a general absence of global age-related methylation patterns (**Figure 6B**). Furthermore, to identify specific CpG sites that could serve as indicators of age, we trained an epigenetic clock with the CG matrix through Lasso Leave-One-Out cross-validation regression (**Figure S10**). However, the correlation between predicted and actual ages was not significant, indicating that age-associated changes in the methylation of CpG sites were not robust enough to build an accurate epigenetic clock.

Next, we explored age-related changes in CpG methylation at the gene level. While meta-gene methylation analysis demonstrated a consistent pattern across age groups (**Figure 6C**), closer examination of the differences between age groups revealed a weak trend of decreasing methylation with advancing age (**Figure S11A**). However, random permutation tests at each position showed no statistically significant differences across all sites for the three pairwise age group comparisons. We also performed linear regression analyses at the individual gene level, using the mean CpG methylation level as the dependent variable and age as the predictor (**Figure S11B**). Among the genes with significant age-associated changes in methylation ($p$-value $< 0.05$), 3,053 genes showed decreased methylation with age, while only 242 genes exhibited increased methylation. Enrichment analysis revealed that the significantly downregulated genes were enriched in pathways related to the modulation of chemical synaptic transmission, hormone metabolic processes, behavioral fear response, and response to xenobiotic stimulus. Conversely, no pathways were enriched among the genes with age-associated increases in methylation. Despite these observations, none of the genes retained significance after p-value adjustment, indicating that gene-level methylation associations with age were not robust.

In addition to gene-level analyses, we performed a differentially methylated regions (DMR) analysis to identify age-related methylation changes. Pairwise comparisons across the three age groups identified 129 significant regions between young and mature, 87 between mature and old, and 84 between

young and old. As in the previous cases, after p-value adjustment, none of these DMRs retained significance.

## DISCUSSION

In this study, our primary focus was elucidating the genome-wide and gene-level methylation patterns in *Daphnia magna* and analyzing the age-associated variations in methylation levels. One key area of interest was the DNA methyltransferases in *D. magna*. Functional annotation via InterProScan revealed that the maintenance methyltransferase *Dma*DNMT1 and de novo methyltransferase *Dma*DNMT3.2 were both expressed at low levels. In contrast, the inactive DNMT3.1 exhibited higher expression levels and a potential association with aging, suggesting functions beyond DNA methylation. For example, DNMT3.1 has been shown to regulate growth and reproduction under starvation conditions (10).

To explore the evolutionary relationships of invertebrate DNA methyltransferases, we compared *Dma*DNMTs with DNMTs from other invertebrates. Oyster *Crassostrea gigas* (mulluscs) has both functional DNMT1 and DNMT3 (36). Among annelid species, *Capitella teleta*, *Ophryotrocha fusiformis*, and *Dinophilus gyrociliatus* possess DNMT1 with active methylase domains. However, as in *D. magna*, both *C. teleta* and *D. gyrociliatus* lack the N-terminal DMAP domain (37), which may point to a shared adaptation in DNMT structure among certain invertebrates. Regarding DNMT3, it is only present in *O. fusiformis* and *C. teleta*. In insects, while the honeybee genome contains active DNMT1 and two DNMT3 variants (Y. Wang et al., 2006), *Drosophila* lacks both DNMT1 and DNMT3 (Krauss & Reuter, 2011). The observed diversity in the presence and functionality of DNMTs among invertebrates reflects adaptive modifications to their unique genomic and environmental demands, with *D. magna* representing one variation of this evolutionary landscape.

We cultured 17 *D. magna* samples of different ages and conducted whole-genome bisulfite sequencing, achieving a high-coverage methylome with an approximate coverage of 60x. This allows us to analyze the methylation patterns across genome, gene, and exon levels. Globally, *D. magna* exhibited an average methylation level of 2%. In comparison, mammals such as mice have an average methylation level of around 80% (38). This difference may be due to the absence of a nuclear localization signal and a DMAP-1 binding domain in *Dma*DNMT1. The nuclear localization signal is crucial for transporting DNMT1 back to the nucleus. The DMAP-1 binding domain facilitates DNMT1's interaction with DMAP-1, a co-repressor that aids its methylation maintenance activity (39). The absence of these features in *D. magna's* DNMT1 may lead to its limited activity and the correspondingly low levels of DNA methylation observed. Additionally, the low expression levels of DNMT1 may further explain the overall low methylation levels.

At the gene level, *D. magna* has an average methylation of around 3%, similar to its Arthropod relative *Apis mellifera* (honey bee) (38). This low level may be attributed to the absence of a functional PWWP domain in DNMT3.2. Since the PWWP domain binds to H3K36me2 and H3K36me3, markers typically found in transcribed regions, its absence prevents the active DNMT3.2 from recognizing this signal in gene bodies, resulting in reduced methylation (40,41).

In examining the methylation patterns across gene bodies and surrounding regions, we observed that the genes are hypomethylated upstream of the TSS and hypermethylated downstream of the TSS, with methylation levels decreasing along the genes. This pattern aligns with findings in the annelids, *O. fusiformis* and *C. teleta*, where a similar trend of hypomethylation at TSS has been reported (37). The hypomethylation at the TSS likely results from the activity of *Dma*DNMT3.2's ADD domain, which selectively binds to unmethylated H3K4 regions (42). As promoters are typically rich in methylated H3K4, DNMT3.2 doesn't bind and thereby prevents DNA methylation (43). This mechanism is supported by a previous study on *D. magna* that documented a high enrichment of H3K4me3 upstream of the actively transcribed *Actin* gene (9).

We observed significantly higher methylation levels in exons compared to introns, which was also reported in *O. fusiformis*, *C. gigas*, and *Apis mellifera* (36–38). This differential methylation is likely influenced by nucleosome positioning, as DNA methyltransferases preferentially target nucleosome-bound DNA (44). Consequently, the enrichment of nucleosomes on exons leads to higher methylation in these regions, suggesting a role for DNA methylation in exon definition and alternative splicing regulation (45,46).

Our analysis of *Daphnia magna* samples revealed minor, insignificant decreases in global and gene-specific methylation levels with age as well as no significant age-related clustering. This contrasts with vertebrates, where DNA methylation patterns are highly dynamic and exhibit notable changes during aging (47). Thus, the relatively constant DNA methylation level of *D. magna* across age underlines the possibility DNA methylation is less coupled to developmental changes in gene expression than in vertebrates, possibly due to the different structure of the DNA methyltransferases. At the same time, our attempt to build an epigenetic clock failed for *D. magna*. Indeed, only a few epigenetics clocks have been reported for invertebrates. Recently, an epigenetic clock on *Nasonia vitripennis*, the jewel wasp, using 19 age-predictive CpG sites was found to have a Spearman's P of 0.94 for the correlation between predicted and actual age (48). However, the fact that the clock CpG sites were pre-selected for their age association indicates possible overfitting to their samples. Thus we still lack robust evidence that epigenetic clocks can be generated for invertebrates.

In this study, our primary emphasis was on the *D. magna* Xinb3 clone. The methylation profiles we elucidated may be slightly altered in other *D. magna* strains. Future research should explore the methylation patterns in additional clones to achieve a more comprehensive understanding. Given the limited correlation between DNA methylation and chronological age, our future attention will shift towards the influence of environmental factors, including temperature and water composition, on methylation dynamics.

**Acknowledgment**

LP was supported by the quadraScope Ventures fund and McKenzie Family Charitable Trust.

**Disclosure of interest**

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

**Data availability of the Statement**

Bisulfite sequencing data and CGmatrix generated in this study have been deposited within the Gene Expression Omnibus (GEO) repository, accession number GSE267859.

## Reference


1. Ebert D. Ecology, epidemiology, and evolution of parasitism in Daphnia. National Library of Medicine; 2005.
2. Tkaczyk A, Bownik A, Dudka J, Kowal K, Ślaska B. Daphnia magna model in the toxicity assessment of pharmaceuticals: A review. Sci Total Environ. 2021;763:143038.
3. Vandegehuchte MB, Lemière F, Vanhaecke L, Berghe WV, Janssen CR. Direct and transgenerational impact on Daphnia magna of chemicals with a known effect on DNA methylation. Comp Biochem Physiol Part C Toxicol Pharmacol. 2010;151(3):278–85.
4. Watson RE, McKim JM, Cockerell GL, Goodman JI. The value of DNA methylation analysis in basic, initial toxicity assessments. Toxicol Sci. 2004;79(1):178–88.
5. Bonini NM, Berger SL. The sustained impact of model organisms—in genetics and epigenetics. Genetics. 2017;205(1):1–4.
6. Hearn J, Plenderleith F, Little TJ. DNA methylation differs extensively between strains of the same geographical origin and changes with age in Daphnia magna. Epigenetics Chromatin. 2021;14:1–14.
7. Smith ZD, Meissner A. DNA methylation: roles in mammalian development. Nat Rev Genet. 2013;14(3):204–20.
8. Kvist J, Gonçalves Athanàsio C, Shams Solari O, Brown JB, Colbourne JK, Pfrender ME, et al. Pattern of DNA methylation in Daphnia: evolutionary perspective. Genome Biol Evol. 2018;10(8):1988–2007.
9. Lindeman LC, Thaulow J, Song Y, Kamstra JH, Xie L, Asselman J, et al. Epigenetic, transcriptional and phenotypic responses in two generations of Daphnia magna exposed to the DNA methylation inhibitor 5-azacytidine. Environ Epigenetics. 2019;5(3):dvz016.
10. Nguyen ND, Matsuura T, Kato Y, Watanabe H. DNMT3. 1 controls trade-offs between growth, reproduction, and life span under starved conditions in Daphnia magna. Sci Rep. 2021;11(1):7326.
11. Lyko F. The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. Nat Rev Genet. 2018;19(2):81–92.
12. Duan R, Fu Q, Sun Y, Li Q. Epigenetic clock: A promising biomarker and practical tool in aging. Ageing Res Rev. 2022;81:101743.
13. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol. 2013;14:1–20.
14. Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nat Rev Genet. 2018;19(6):371–84.
15. Klüttgen B, Dülmer U, Engels M, Ratte H. ADaM, an artificial freshwater for the culture of zooplankton. Water Res. 1994;28(3):743–6.
16. Cho Y, Jonas-Closs RA, Yampolsky LY, Kirschner MW, Peshkin L. Intelligent high-throughput intervention testing platform in Daphnia. Aging Cell. 2022;21(3):e13571.
17. Fields PD, Obbard DJ, McTaggart SJ, Galimov Y, Little TJ, Ebert D. Mitogenome phylogeographic analysis of a planktonic crustacean. Mol Phylogenet Evol. 2018;129:138–48.

18. Farrell C, Thompson M, Tosevska A, Oyetunde A, Pellegrini M. BiSulfite Bolt: A bisulfite sequencing analysis platform. Gigascience. 2021;10(5):giab033.
19. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10(2):giab008.
20. Guo W, Zhu P, Pellegrini M, Zhang MQ, Wang X, Ni Z. CGmapTools improves the precision of heterozygous SNV calls and supports allele-specific methylation detection and visualization in bisulfite-sequencing data. Bioinformatics. 2018;34(3):381–7.
21. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics J Integr Biol. 2012;16(5):284–7.
22. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14(2):178–92.
23. McInnes L, Healy J, Melville J. Umap: Uniform manifold approximation and projection for dimension reduction. ArXiv Prepr ArXiv180203426. 2018;
24. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res. 2011;12:2825–30.
25. Jühling F, Kretzmer H, Bernhart SH, Otto C, Stadler PF, Hoffmann S. Metilene: fast and sensitive calling of differentially methylated regions from bisulfite sequencing data. Genome Res. 2016;26(2):256–62.
26. Dua I, Pearson AC, Lowman RL, Peshkin L, Yampolsky LY. Post-senescence reproductive rebound in Daphnia associated with reversal of age-related transcriptional changes. GeroScience. 2024;1–20.
27. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:1–21.
28. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30(9):1236–40.
29. Genereux DP, Johnson WC, Burden AF, Stöger R, Laird CD. Errors in the bisulfite conversion of DNA: modulating inappropriate-and failed-conversion frequencies. Nucleic Acids Res. 2008;36(22):e150–e150.
30. Bird AP. Use of restriction enzymes to study eukaryotic DNA methylation: II. The symmetry of methylated sites supports semi-conservative copying of the methylation pattern. J Mol Biol. 1978;118(1):49–60.
31. Mechta M, Ingerslev LR, Fabre O, Picard M, Barrès R. Evidence suggesting absence of mitochondrial DNA methylation. Front Genet. 2017;8:166.
32. Bensasson D, Zhang DX, Hartl DL, Hewitt GM. Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends Ecol Evol. 2001;16(6):314–21.
33. Kowal K, Tkaczyk A, Pierzchała M, Bownik A, Ślaska B. Identification of mitochondrial dna (Numts) in the nuclear genome of daphnia magna. Int J Mol Sci. 2020;21(22):8725.
34. Moore LD, Le T, Fan G. DNA methylation and its basic function. Neuropsychopharmacology. 2013;38(1):23–38.

35. Suzuki MM, Kerr AR, De Sousa D, Bird A. CpG methylation is targeted to transcription units in an invertebrate genome. Genome Res. 2007;17(5):625–31.
36. Wang X, Li Q, Lian J, Li L, Jin L, Cai H, et al. 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. 2014;15:1–12.
37. Guynes K, Sarre LA, Carrillo-Baltodano AM, Davies BE, Xu L, Liang Y, et al. Annelid methylomes reveal ancestral developmental and aging-associated epigenetic erosion across Bilateria. Genome Biol. 2024;25(1):204.
38. Feng S, Cokus SJ, Zhang X, Chen PY, Bostick M, Goll MG, et al. Conservation and divergence of methylation patterning in plants and animals. Proc Natl Acad Sci. 2010;107(19):8689–94.
39. Lee GE, Kim JH, Taylor M, Muller MT. DNA methyltransferase 1-associated protein (DMAP1) is a co-repressor that stimulates DNA methylation globally and locally at sites of double strand break repair. J Biol Chem. 2010;285(48):37630–40.
40. Dukatz M, Holzer K, Choudalakis M, Emperle M, Lungu C, Bashtrykov P, et al. H3K36me2/3 binding and DNA binding of the DNA methyltransferase DNMT3A PWWP domain both contribute to its chromatin interaction. J Mol Biol. 2019;431(24):5063–74.
41. Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS, Ahringer J. Differential chromatin marking of introns and expressed exons by H3K36me3. Nat Genet. 2009;41(3):376–81.
42. Otani J, Nankumo T, Arita K, Inamoto S, Ariyoshi M, Shirakawa M. Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX–DNMT3–DNMT3L domain. EMBO Rep. 2009;10(11):1235–41.
43. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell. 2006;125(2):315–26.
44. Chodavarapu RK, Feng S, Bernatavichute YV, Chen PY, Stroud H, Yu Y, et al. Relationship between nucleosome positioning and DNA methylation. Nature. 2010;466(7304):388–92.
45. Ehrlich M, Lacey M. DNA methylation and differentiation: silencing, upregulation and modulation of gene expression. Epigenomics. 2013;5(5):553–68.
46. Shayevitch R, Askayo D, Keydar I, Ast G. The importance of DNA methylation of exons on alternative splicing. Rna. 2018;24(10):1351–62.
47. Meehan RR, Stancheva I. DNA methylation and control of gene expression in vertebrate development. Essays Biochem. 2001;37:59–70.
48. Brink K, Thomas C, Jones A, Chan T, Mallon E. Exploring the ageing methylome in the model insect, Nasonia vitripennis. BMC Genomics. 2024;25(1):305.

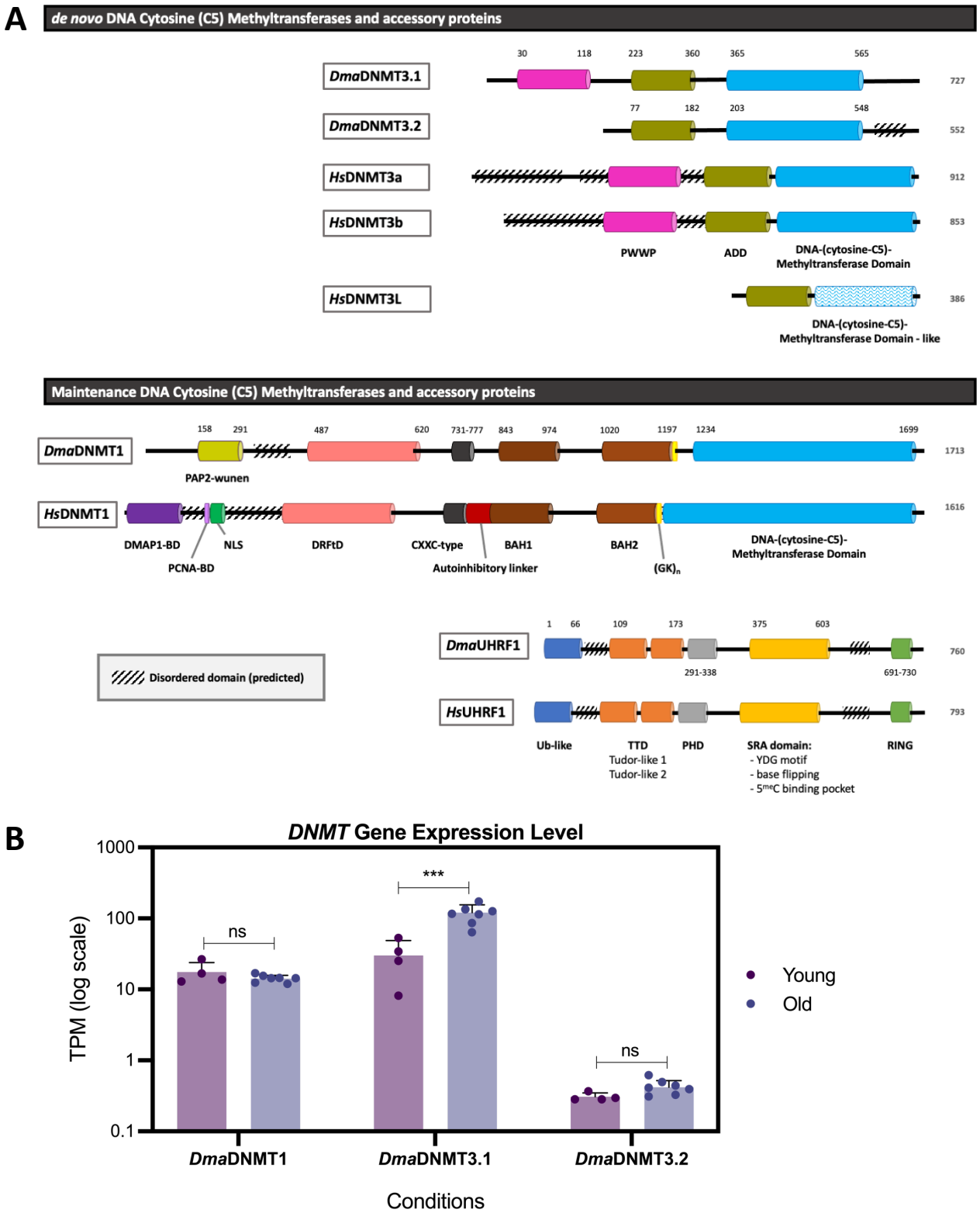


**Figure 1. The comparison of DNA (C5) methyltransferases and accessory proteins of *D. magna* and *H. sapiens* and the gene expression levels under different conditions. (A)** Domain structures of *de novo* and maintenance DNA methyltransferases and accessory proteins in *Daphnia magna* and *Homo sapiens* were predicted using InterProScan. Colored boxes represent the functional domains, with numbers indicating their respective positions along the protein sequence. Diagonal black-striped boxes indicate a predicted disordered consensus sequence. **(B)** Gene expression levels (TPM on log scale) of three DNMT genes (*Dma*DNMT1, *Dma*DNMT3.1, and *Dma*DNMT3.2) in young and old *D. magna* individuals. Bars represent the mean expression levels, and error bars denote standard deviation (SD). Statistical significance between conditions for each gene was assessed using a two-sided Student's t-test, with significance indicated as follows: $P < 0.001$ (***) and nonsignificant (ns). Expression levels of *Dma*DNMT1 and *Dma*DNMT3.2 showed no significant differences between the groups, while *Dma*DNMT3.1 expression was significantly different in young individuals compared to the old group.

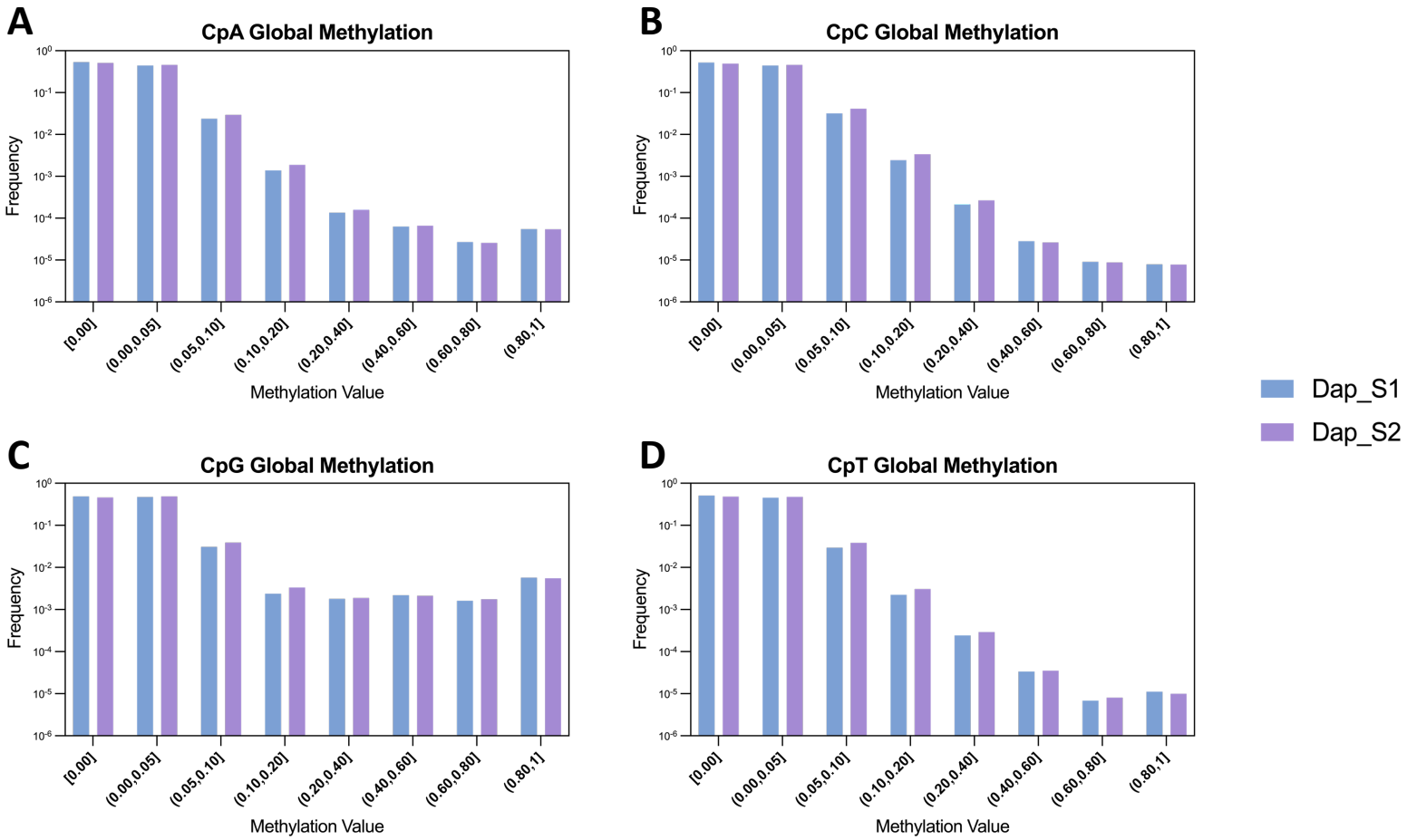


**Figure 2. Global DNA methylation levels at CpN sites of two deeply sequenced samples.** Methylation levels of individual CpN sites are presented as the frequency distribution. **(A)** CpA, **(B)** CpC, and **(D)** CpT sites comparably exhibit low methylation levels. **(C)** CpG sites display a significantly higher degree of methylation overall.

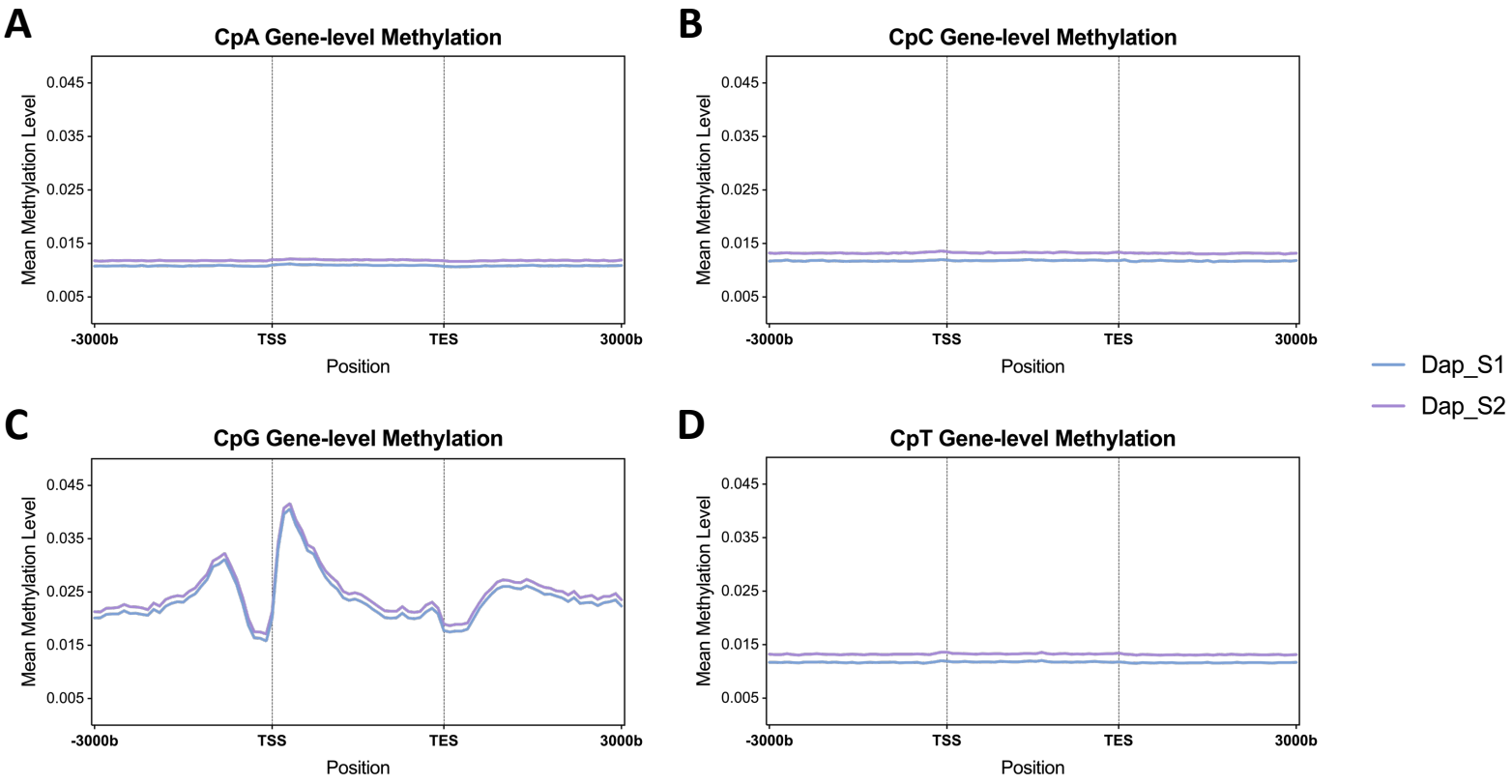


**Figure 3. Gene-level DNA CpN methylation levels in two deeply sequenced samples, spanning 3 kilobases (kb) upstream transcription start sites (TSS) and 3 kb downstream of transcription end sites (TES).** Each line depicts the average gene methylation value for **(A)** CpA, **(B)** CpC, **(C)** CpG, and **(D)** CpT for a sample. Non-CpG sites exhibited uniformly low methylation levels, whereas CpG sites manifest a distinctive pattern. These sites have lower methylation levels at TSS compared to regions upstream and downstream. TSS: transcription start site; TES: transcription end site.

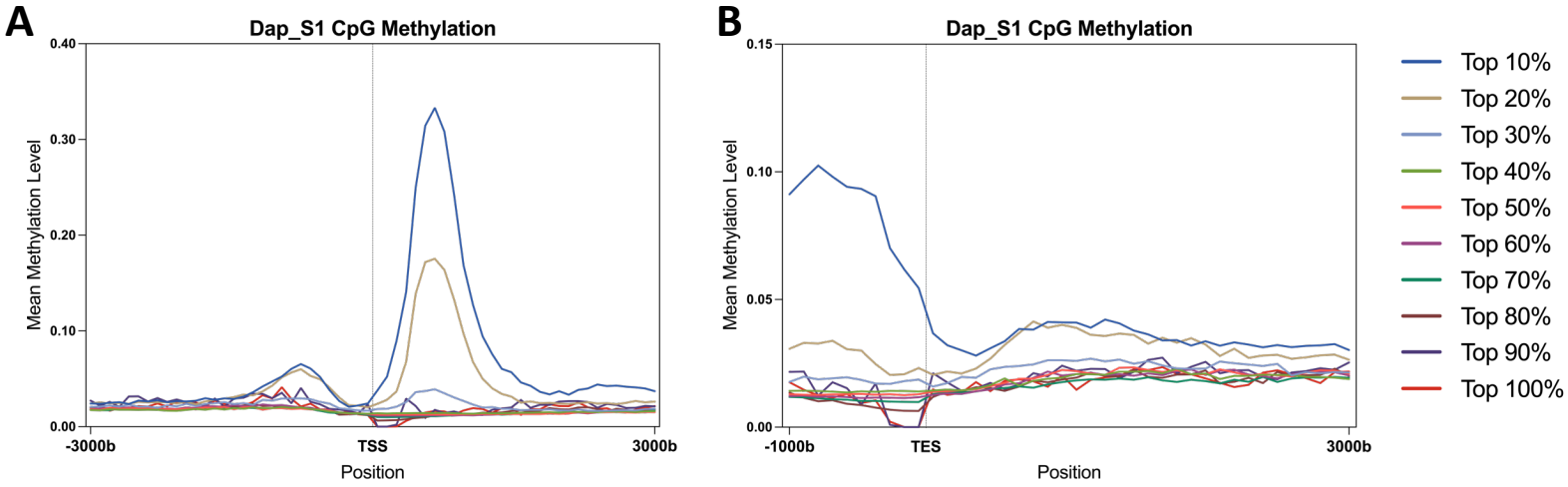


**Figure 4. Gene-level CpG Methylation Patterns Centered around TSS and TES for Dap_S1.** Genes were divided into 10 groups based on their mean methylation levels. Each line represents the average methylation level within each group, centered around **(A)** the transcription start site (TSS) and **(B)** the transcription end site (TES). For example, Top 20% in legend represents the top 10-20% group. TSS: transcription start site; TES: transcription end site.

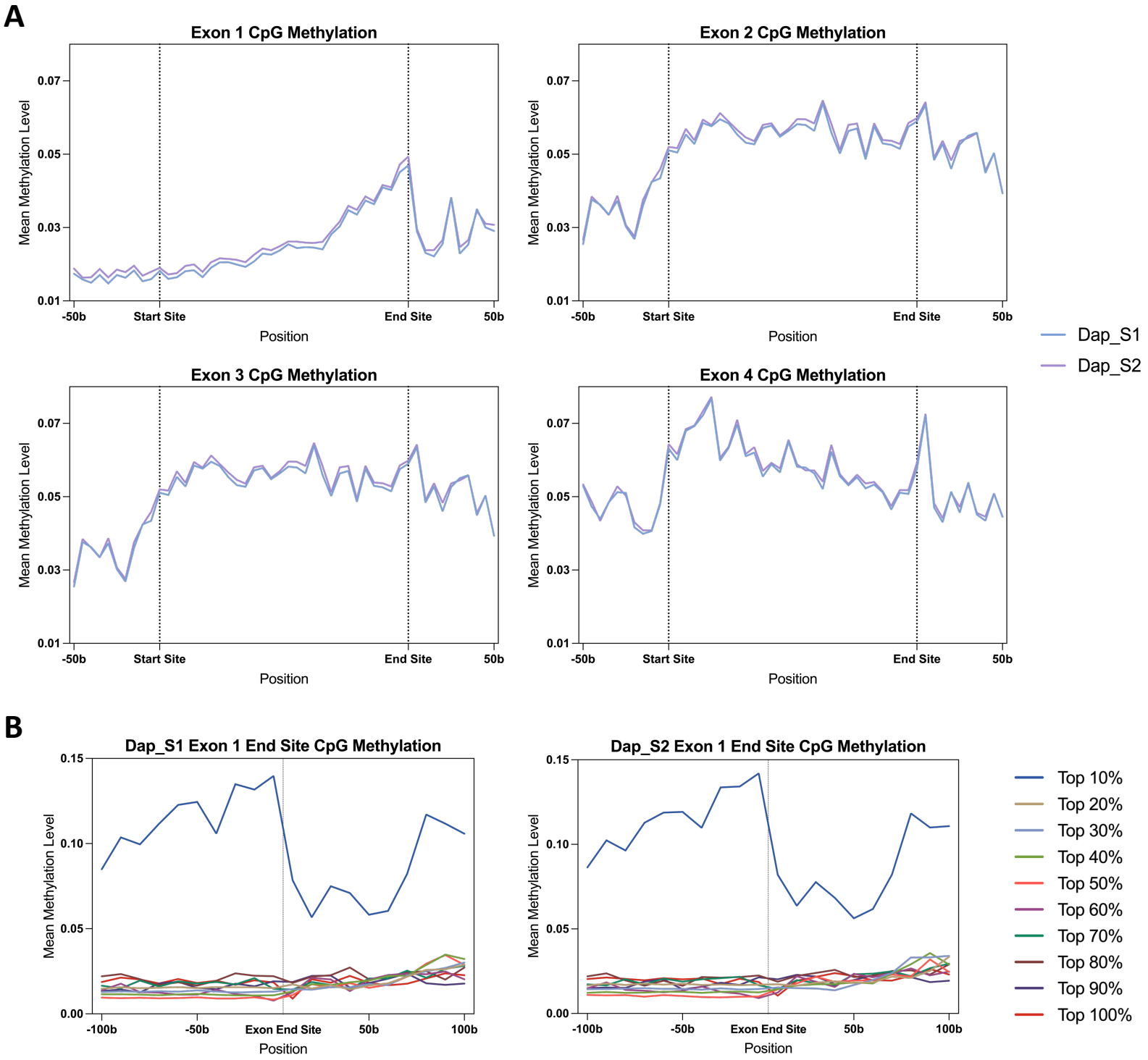


**Figure 5. Exon-level CpG Methylation Patterns in two deeply sequenced samples. (A)** CpG methylation levels of Exon 1 to Exon 4 from genes with more than 5 exons are demonstrated. Exon 1 shows a gradual increase in CpG methylation, peaking near its 3' end. Methylation levels continued to rise from Exon 2 through Exon 4, reaching a peak of nearly 8% in Exon 4 before decreasing. **(B)** Genes were divided into 10 groups based on their mean methylation levels in Exon 1. Each line represents the average methylation level within each group, centered around Exon 1 end site, with 100b up- and downstream of each region. In top 10% genes, there is a sharp drop in methylation transitioning from Exon 1 to the adjacent intron.

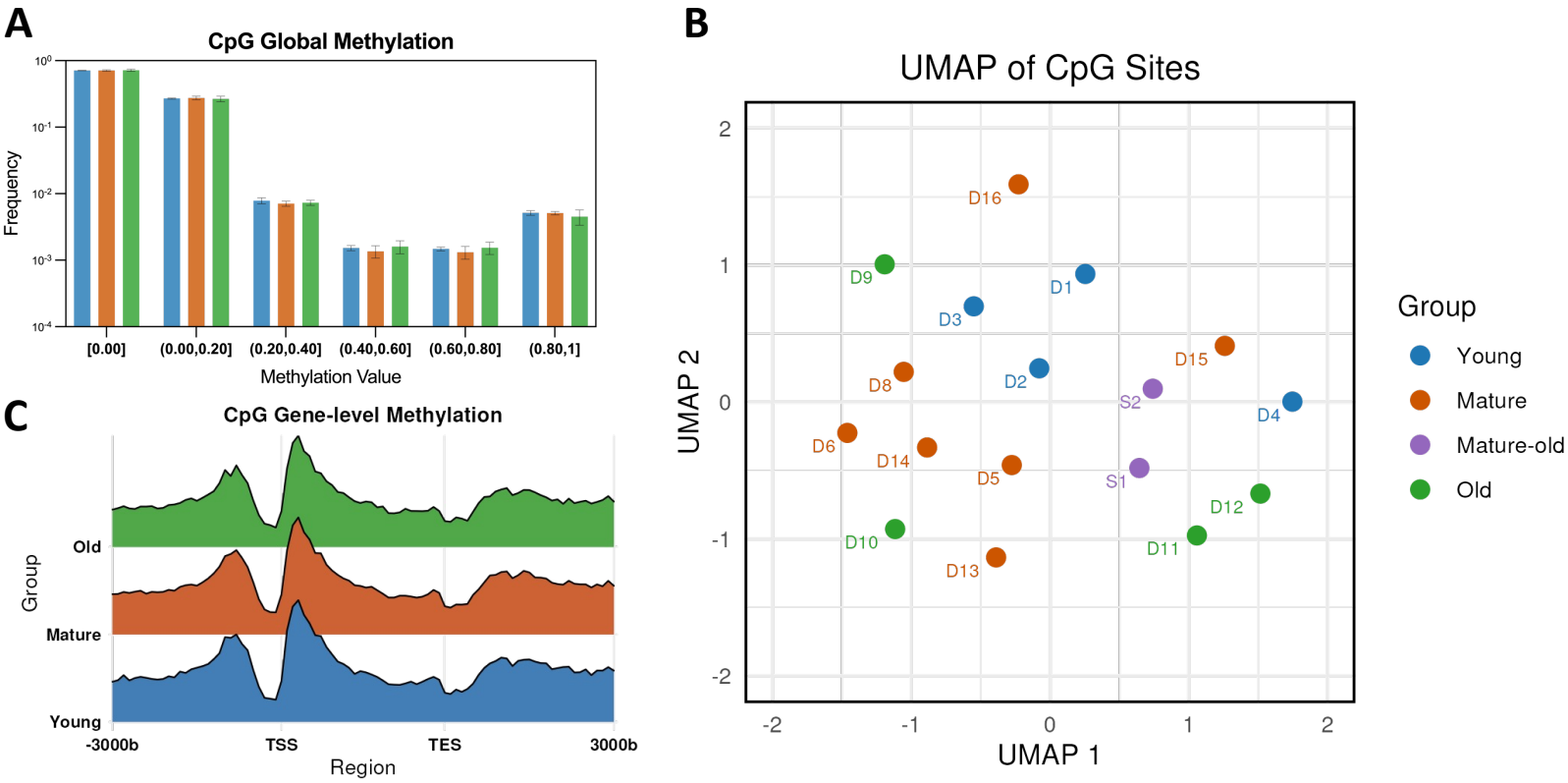


**Figure 6. CpG methylation patterns across age groups demonstrate weak relation to age. (A)** Methylation levels of individual CpG sites in each group are presented as the frequency distribution. The decrease in average methylation levels correlate an increase in age. **(B)** The UMAP visualization of matrix with CpG sites present in at least 80% of the 17 samples does not reveal any specific age-related clustering. **(C)** Meta-gene ridge plots plot across age groups demonstrate a similar pattern.