Global Ecology and Conservation 69 (2026) e04284

Contents lists available at ScienceDirect

Global Ecology and Conservation

journal homepage: www.elsevier.com/locate/gecco

Original research article

Characteristics of fecal DNA methylation and usability in tiger age estimation

Wen Hui Wang [a,b], Qi Zhang [a,b], Jia Ning Chu [a,b], Xiao Xi Liu [c], Peng Gao [a,b], Ao Zou [a,b], Dan Liu [b,d], Hai Tao Xu [d], Hai Tao Huang [d], Shu Hui Yang [a,b,*], Yan Chun Xu [a,b,*]

[a] College of Wildlife and Protected Area, Northeast Forestry University, Harbin, China
[b] National Forestry and Grassland Administration Research Center of Engineering Technology for Wildlife Conservation and Utilization, Harbin, China
[c] School of Civil Engineering and Transportation, Northeast Forestry University, Harbin, China
[d] Heilongjiang Siberian Tiger Park, Harbin, China

ARTICLE INFO

Keywords:
Amur tiger
DNA methylation
DNA methylation heterogeneity
Age estimation

ABSTRACT

Age structure is a key determinant of population dynamics and conservation management in wildlife, yet reliable age estimation typically depends on invasive sampling. Epigenetic clocks enable DNA-based age prediction across taxa, but their application to wildlife remains limited by reliance on blood or tissue samples. Feces offer a non-invasive alternative, although fecal-based age prediction has so far shown limited accuracy. Using captive Amur tigers (*Panthera tigris altaica*) of known ages as a model, we characterized DNA methylation patterns in fecal nuclear DNA, primarily derived from exfoliated intestinal epithelial cells, via whole-genome bisulfite sequencing. We identified 512 age-associated methylation sites and constructed age prediction models using Elastic Net regression, Support Vector Machine, and Random Forest regression based on methylation levels and methylation heterogeneity. The Random Forest model achieved the best performance, with a coefficient of determination (R^2) of 0.981 and a mean absolute error (MAE) of 7.18 months in the training set, and an R^2 of 0.972 with an MAE of 7.37 months in the test set. Prediction error accounted for approximately 3.35% of the maximum sampled age (220 months) and 2.46–2.79% of the known lifespan of captive Amur tigers, outperforming previously reported fecal- or tissue-based epigenetic clocks in mammals. These results demonstrate that fecal nuclear DNA retains sufficient age-associated epigenetic information for accurate age estimation. We further propose a captive-to-wild equivalent age framework to facilitate application in free-ranging populations. Despite of the current constraints of sample size and validation strategy, this preliminary and exploratory study suggests the utility of feces and technical framwork for age structure inference in wildlife management and conservation practice.

1. Introduction

Age structure is a crucial parameter related to the growth potential of a population, and estimating the age of individuals is

* Corresponding authors at: College of Wildlife and Protected Area, Northeast Forestry University, Harbin, China.
E-mail addresses: 474568251@qq.com (S.H. Yang), xu_daniel@163.com (Y.C. Xu).

https://doi.org/10.1016/j.gecco.2026.e04284
Received 16 January 2026; Received in revised form 3 June 2026; Accepted 3 June 2026
Available online 4 June 2026

therefore fundamental for understanding population dynamics and informing wildlife conservation. Traditionally, age estimation in animals has relied largely on morphological or anatomical traits that change predictably with age. For example, growth rings in tooth cross-sections (Wehausen et al., 2024) and bones (Wei et al., 1988) can record seasonal dynamics in tissue growth, whereas the degree of tooth or bone wear reflects the duration and intensity of use (Milner-Gulland and Akçakaya, 2001; Stander, 2015). Morphometric indices of skeletal and body growth also change gradually with age and have been used for age estimation in some taxa (Armstrong and Brooks, 2014). These approaches are commonly applied to mammals, particularly carnivores and ungulates, while other taxa rely on different age-informative structures, such as otoliths in fish species (Pacheco et al., 2021).

Despite their utility, these classical age-estimation methods often require hard tissues or direct physical measurements, which may involve invasive procedures, carcass collection, or even lethal sampling. Such requirements limit their applicability to free-ranging wildlife, especially rare and endangered species for which repeated capture or destructive sampling is ethically and practically unacceptable. Consequently, there is a strong need for age-estimation techniques based on minimally invasive or non-invasive materials that can be collected safely and repeatedly in wildlife research, management, and conservation.

To overcome these limitations, several alternative age-estimation approaches have recently been developed based on physiological, biochemical, and genomic parameters. These include sex hormone levels (Cattet et al., 2018), chromosomal telomere length (Jansen et al., 2021), transcriptomic profiles (Holly et al., 2013; Peters et al., 2015), proteomic signatures (Krištić et al., 2014; Menni et al., 2015), metabolomic markers (Mutz and Iniesta, 2024), and DNA methylation patterns (Lowe et al., 2020). Among these approaches, epigenetic clocks based on age-associated DNA methylation have attracted increasing attention because methylation changes at specific CpG sites can provide reproducible molecular signatures of aging. However, many molecular approaches still depend on tissues such as blood, skin, or internal organs, which remain difficult to obtain from wild animals. Therefore, developing epigenetic clocks from non-invasive materials represents a particularly promising direction for wildlife age estimation.

Feces are among the most accessible non-invasive biomaterials for wildlife studies, as they contain both gut microbial DNA and host DNA derived from exfoliated intestinal epithelial cells. Although microbial community shifts have enabled the development of "microbiome aging clocks" in humans (Galkin et al., 2020; Ira et al., 2024), their application in wildlife remains challenging. For instance, age inference in Amur tigers is currently limited to broad classes (Hu et al., 2025). This limitation is exacerbated by the microbiome's sensitivity to host diet, seasonal fluctuations (Ma, 2024; Neha and Salazar-Bravo, 2023; Sun et al., 2021), and rapid post-defecation degradation (Hale et al., 2016), which collectively compromise the reliability of microbiome-based aging markers.

To overcome these constraints, targeting host DNA derived from fecal epithelial cells provides a more robust, tissue-specific alternative (Cui et al., 2024; Cuomo et al., 2023). Host aging is intrinsically linked to epigenetic modifications, particularly DNA methylation at CpG sites. These sites follow two distinct aging patterns: 'drifting,' which involves a stochastic increase in epigenetic disorder, and 'steady,' characterized by directional gains or losses of methylation over time (Moqri and Poganik, 2025). Since these age-dependent methylation changes are systemic and reproducible, they enable the construction of epigenetic clocks to predict biological age and health status (Duan et al., 2022). Therefore, leveraging host DNA methylation within fecal samples offers a more stable and accurate approach for monitoring age in wildlife populations compared to the highly variable microbial community.

Several epigenetic clocks have been proposed, ranging from multi-tissue estimators like the Horvath clock (Horvath, 2013) to tissue-specific models like the Hannum clock (Hannum et al., 2013). These studies demonstrate the utility of epigenetic modifications in predicting biological age and provide viable frameworks for clock development. A number of models have been established for various species, including naked mole-rats (*Heterocephalus glaber*; liver/skin) (Lowe et al., 2020), domestic dogs (*Canis lupus familiaris*) and wolves (*Canis lupus*; blood) (Thompson et al., 2017), domestic cats (*Felis catus*; blood) (Qi et al., 2021), and Bechstein's bats (*Myotis bechsteinii*; wing tissue) (Wilkinson et al., 2021; Wright and Mathews, 2018). Given the practical advantages of non-invasive sampling in wildlife, recent efforts have extended to fecal epigenetic clocks in Indo-Pacific bottlenose dolphins (*Tursiops aduncus*) (Yagi, Qi, 2024) and mice (*Mus musculus*) (Hanski et al., 2024). However, these approaches often rely on methylation markers previously identified in blood or tissue, which remains a significant constraint due to the limited availability of such reference data for most wildlife. Nevertheless, these studies underscore that fecal DNA from intestinal epithelial cells retains biologically meaningful, age-associated methylation signals, representing a promising substrate for further epigenetic clock development.

The Amur tiger (*Panthera tigris altaica*) is an apex predator and one of the world's most endangered large carnivores (Wen et al., 2022), primarily inhabiting northeastern China and the Russian Far East. Extensive conservation efforts have been directed toward this subspecies, covering habitat suitability (Li et al., 2017, 2016), spatial distribution and population dynamics (Shaochun et al., 2008), movement and corridor establishment (Wang et al., 2023), evolutionary history and genetic diversity (Luo et al., 2019), and law enforcement (Xu et al., 2005). Despite these efforts, the age structure of wild populations remains largely unknown. Consequently, assessing population growth potential from an age-specific dimension has remained elusive, highlighting a critical gap in our ability to monitor and manage this species effectively.

Feces is the most conveniently accessible sample for this large carnivore both in the wild and in captivity. In this study, we isolated and enriched host DNA from fecal samples and performed Whole-Genome Bisulfite Sequencing (WGBS) to identify methylated sites. We further characterized methylation patterns, haplotype blocks, and single-molecule methylation heterogeneity to develop robust epigenetic clocks for age estimation. Although this study represents a preliminary exploration of non-invasive epigenetic aging, our findings demonstrate the feasibility of using fecal WGBS data for age prediction, despite the inherent challenges of lower sequencing coverage compared to traditional tissue or blood samples (Yang JC et al., 2026). An epigenetic clock was successfully developed for captive tigers with known chronological ages. The proposed approach not only provides a non-invasive tool for age estimation in captive tigers but also offers a potential framework, foundational methodology and valuable experience that could facilitate the development of de novo epigenetic clocks for a broader range of wildlife species.

2. Materials and methods

2.1. Ethics statement

This study was approved by the Ethics Committee of Northeast Forestry University (No. 20200512 on May 12, 2020). All procedures were performed in accordance with the ethical guidelines of the 1975 Declaration of Helsinki.

2.2. Sample collection and fecal DNA extraction

A total of 40 fecal samples were initially collected from Amur tigers aged from 10 to 220 months at the Heilongjiang Siberian Tiger Park (Harbin, China) and the China Hengdaohezi Feline Breeding Centre (Hengdaohezi town, Mudanjiang, China). Individual identifiers and information regarding birth date, sex, and sampling date were clearly documented. The samples were all fresh at the time of collection and immediately stored at −80°C. Total DNA was extracted from the fecal samples with enrichment of the host fraction following the Peri-extraction Enrichment by SDS (PEERS) protocol established in Cui et al. (2024), see Supplementary Method S1 for detailed operational procedures and specific parameters).

To ensure data reliability, we utilized the endogenous nuclear DNA enrichment efficiency (EE_N) as a quality control metric. EE_N is defined as the ratio of the copy number of endogenous nuclear DNA (CN_N) to the copy number of bacterial genomic DNA (CN_B) (Cui et al., 2024). Samples were required to meet a minimum EE_N threshold of 1×10^{-3}, a level previously demonstrated to yield acceptable mapping rates, genome coverage, and sequencing depth. Ultimately, 33 samples (82.5%) successfully met this criterion and were selected for subsequent bisulfite conversion and downstream epigenetic analyses, while 7 samples failed to reach the 1×10^{-3} threshold and were excluded.

2.3. DNA conversion, sequencing, and bioinformatics analysis

2.3.1. DNA conversion and Sequencing

Each DNA was treated with bisulfite using the EZ DNA Methylation Kit (ZYMO Research, Irvine, USA) to convert unmethylated cytosines to uracil while keeping methylated cytosines unaltered. The library was then prepared using the xGen™ Methyl-Seq DNA Library Prep Kit (IDT, USA) and sequenced on an Illumina HiSeq 2000 platform (Illumina, USA) to a depth of 80 ×.

2.4. Bioinformatics analysis

Each fragment produced 150 bp paired-end reads. The raw paired-end sequencing data were processed using Trim Galore v0.6.7 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) to remove technical artifacts. Specifically, explicit Illumina adapter sequences were trimmed first. To eliminate positional sequencing biases, hard clipping was applied to the read ends: 10 bp and 15 bp were removed from the 5′ ends of the forward and reverse reads, respectively, while 10 bp were truncated from the 3′ ends of both reads. Subsequently, low-quality terminal bases (Phred score < 20) were trimmed, and paired reads where either read became shorter than 20 bp were discarded from downstream analysis. The fastp v0.23.0 (Chen et al., 2018) program was used for quality control. Clean reads were mapped to the wild Amur tiger reference genome (PtaHapG) (Lan et al., 2025) with BSMAP v2.42 (Xi and Li, 2009) with 3 threads, allowing up to one mismatch (-v 1) and up to eight gaps (-m 8). To examine whether prey-derived DNA contamination was present in the tiger sequencing reads, we performed an additional screening by mapping all reads against a reference panel of potential prey species. Reads that uniquely aligned to non-tiger genomes were quantified to assess the extent of exogenous DNA contamination. Then, MethylDackel v0.6.0 (available online: https://github.com/dpryan79/MethylDackel?tab=readme-ov-file#citing-methyldackel) was used to extract methylation rates report for each sample. This analysis generated coverage statistics for methylated cytosine reads and their contexts (CG, CHG and CHH, where H can be A, T, C, respectively). Metheor v0.1.8 (Lee et al., 2023) was applied to extract each sample's DNA methylation heterogeneity metrics (Scherer et al., 2020), including methylation entropy (ME), proportion of discordant reads (PDR), epipolymorphism (PM), fraction of discordant read pairs (FDRP), and quantitative fraction of discordant read pairs (qFDRP). All sites measured at least five times were retained. Sites with excessively high coverage, potentially reflecting PCR amplification bias, as well as sites with high proportions of missing values across samples, were excluded.

2.5. Assessment of host DNA enrichment and WGBS data yield

To assess the impact of host nuclear DNA enrichment efficiency (EE_N) on whole-genome bisulfite sequencing (WGBS) data yield, we analyzed the relationships between EE_N and two sequencing output metrics: the number of valid CpG sites with a sequencing depth of at least 5 × and the genome-wide host DNA coverage. Following alignment to the host reference genome and standard quality filtering, effective CpG sites and genome coverage were quantified for each sample. The associations between EE_N and sequencing output were modeled using logistic functions to characterize saturation behavior, with the upper asymptote (K) representing the theoretical maximum and the EE_N value corresponding to 95% of K defined as the saturation threshold.

2.6. Subsampling-based evaluation of WGBS sequencing depth

To assess whether sequencing depth was sufficient, we performed proportional subsampling of raw whole-genome bisulfite sequencing (WGBS) reads for five individuals (10–100% of total reads in 10% increments). At each subsampling level, the number of detected CpG sites was quantified. The relationship between sequencing depth and CpG site detection was modeled independently for each individual using a logistic function, with the plateau parameter representing the theoretical maximum number of detectable CpG sites. Sequencing depth adequacy was evaluated by estimating the depth required to reach 95% of the plateau, defined as the point beyond which further increases in sequencing depth resulted in minimal gains in detected CpG sites. Uncertainty in the estimated 95% plateau depth was assessed using nonparametric bootstrap resampling (500 replicates), and results are reported as the bootstrap mean with 95% confidence intervals.

2.7. Global methylation landscape of nuclear genome

The distribution of DNA methylation and its heterogeneity was examined on nuclear genome using Origin v.2024. Out of the 33 total samples, 15 individuals were selected to represent three distinct life stages ($n = 5$ per group) for comparative analysis: juvenile (10–37 months), adult (88–100 months), and elderly (137–220 months) (see in the Figure S1). This tripartite grouping was specifically designed to identify potential age-related differentially methylated regions (DMRs) and to explore the fundamental patterns of epigenetic heterogeneity across the lifespan. This exploratory analysis served as a foundational step for the subsequent construction of a more precise, continuous epigenetic age clock using the full cohort ($N = 33$). Kruskal-Wallis variance analysis along with Dunn's test was employed to test the differences of both DNA methylation rate and heterogeneity among three age groups. To analyze differentially methylated genes (DMGs) and differentially methylated regions (DMRs), the juvenile, adult and elderly group were compared using DSS v.2.44.0 (Feng et al., 2014) to deal with each sample's methylation report with the criteria Q-value < 0.01 to quantify DMRs. DMRs with an absolute value of diff.Methy≥ 0.30 were identified as focal differentially methylated loci. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was used to determine the roles of DMGs. Pathway data were extracted from standard public pathway databases KEGG (http://www.kegg.jp/kegg/pathway.html). The KEGG pathways were considered significantly enriched with unadjusted raw $P < 0.05$ and ranked by $-\log_{10}(P)$ from lowest to highest.

2.8. Building and assessing the age prediction models

2.8.1. Data integration

All analyses were performed in R v4.1.2. Cytosine-level DNA methylation and heterogeneity metrics (ME, PM, FDRP and qFDRP) were first integrated into a unified dataset using the *dplyr* package (available at https://CRAN.R-project.org/package=dplyr). Rows containing null values were removed from the dataset. Prior to selection of age-associated sites, CpG methylation rate and heterogeneity metrices were integrated, subjected to filtering, and normalized.

2.9. Selection of age-associated sites

To identify age-associated CpG sites and heterogeneity metrics, we performed a genome-wide Spearman correlation analysis between each site and chronological age. CpG sites with strong correlations ($|\rho| \geq 0.5$) were retained as candidate features, representing sites whose methylation or heterogeneity levels increased or decreased with age. This conservative threshold ensured robust identification of biologically meaningful, age-associated loci.

2.10. Epigenetic clock construction and model optimization

2.10.1. Data partitioning and feature selection

The dataset (total $N = 33$) was randomly partitioned into an approximate 80% training set ($n = 26$) and a 20% independent test set ($n = 6$) prior to feature selection and model training. To determine the optimal number of loci for model construction, we calculated the Spearman rank correlation coefficient (ρ) between each methylation or methylation-heterogeneity site and chronological age in the training set. This non-parametric measure captures monotonic associations and is robust to non-normal methylation distributions. All loci were ranked by the absolute correlation value ($|\rho|$), and feature subsets were generated by selecting the top N loci (where N ranged from 200 to 800, e.g., 200, 400, 600 and 800). Each subset was then used for model training to assess how feature dimensionality affects prediction accuracy. This correlation-based filtering strategy is model-independent and fully deterministic, ensuring unbiased feature evaluation and reproducible results.

2.10.2. Model training and model evaluation

Using the selected CpG features, three supervised learning models were trained: Elastic Net regression (*glmnet*), Support Vector Regression with a radial basis function kernel (SVM; *e1071*), and Random Forest (RF; *randomForest*). Model performance was evaluated using leave-one-out cross-validation (LOOCV), implemented in a nested framework to prevent information leakage. For each LOOCV iteration, all feature selection, data preprocessing, and hyperparameter optimization were performed exclusively within the training set, and predictions were generated only for the held-out sample. Predictor variables were standardized (centered and scaled) using parameters estimated from the training data and subsequently applied to the test sample. Hyperparameters were optimized by grid

search within the training folds: alpha values ranging from 0 to 1 and lambda values from 0.001 to 0.1 for the elastic net; sigma values from 0.001 to 0.1 and cost (C) values from 2^2 to 2^9 for the SVM; and mtry values spanning the number of selected CpG features for the random forest.

Model performance was assessed in both training and independent test sets using mean absolute error (MAE), root mean squared error (RMSE), and coefficient of determination (R^2). Scatterplots of predicted versus chronological age were generated to evaluate calibration.

2.10.3. Clock validation

The resulting ensemble-based epigenetic clock was evaluated for its correlation with chronological age and its predictive error (MAE) on the independent test set. Additional analyses can incorporate mixed-effects modeling to quantify the influence of biological sex, sequencing batch, or repeated measures on epigenetic age estimates.

2.10.4. Influence of sequencing depth on model performance

Seqkit(Shen et al., 2016) was utilized to randomly subsample reads from the cleaned raw data of each sample to create subdatasets with average sequencing depths of 5 × , 20 × , 40 × , and 60 × . Each subdataset was then subjected to analysis from step 5.1 through 5.3.2.

2.10.5. KEGG pathway enrichment of the age-associated sites

Gene enrichment analysis was conducted on the genomic sites showing significant age correlation using the ClusterProfiler package (Yu et al., 2012) and the KEGG database (http://www.kegg.jp/kegg/pathway.html).

3. Results

3.1. Data quality of whole genome bisulfite sequencing

Whole genome bisulfite sequencing (WGBS) was successfully conducted on all 33 fecal samples of Amur tigers. On average, each sample yielded 1.44×10^9 clean reads, with a range spanning from 8.57×10^8–2.79×10^9 ones. The average base call accuracy was 91.84% at quality score of Q20, with values ranging from 87.66% to 97.89%, and 74.52% at Q30 with a range of 71.11–94.46%. The average effective read length ratio was 89.27% ranging from 87.51% to 98.09% (refer to Supplementary materials Table S1). The 95% confidence interval of genome coverage was [77.29%, 90.17%], with the average coverage of 83.73% (refer to Supplementary materials Table S2). After reads were stringently aligned to the Amur tiger reference genome using BSMAP, further filtering detected no prey-derived DNA.

3.2. Evaluation of host DNA enrichment and sequencing depth adequacy

Host DNA enrichment efficiency (EE_N) was strongly associated with WGBS data yield and exhibited clear saturation behavior. The

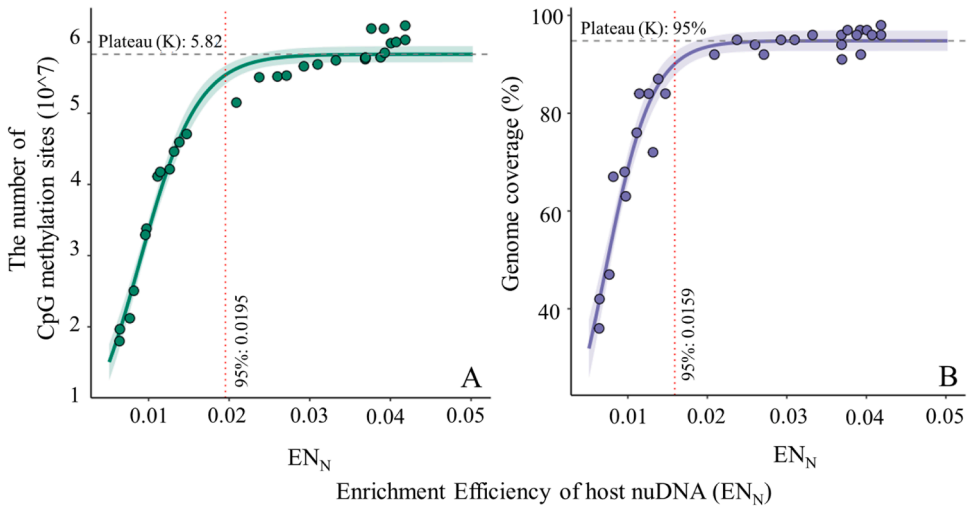

Fig. 1. Relationship between host nuclear DNA enrichment efficiency (EE_N) and whole-genome bisulfite sequencing (WGBS) data yield from fecal DNA. Fig. 1(A) shows the association between EE_N and the number of CpG sites with non-zero coverage, while the Fig. 1(B) shows the relationship between EE_N and genome-wide host DNA coverage. Solid curves represent fitted logistic models, and shaded areas indicate 95% confidence intervals. Horizontal dashed lines denote the estimated saturation plateaus (K), and vertical red dashed lines indicate the EE_N values at which sequencing output reaches 95% of the corresponding plateau. Points represent individual samples.

number of effective CpG sites increased rapidly at low EE_N values and approached an asymptote at higher enrichment levels (Fig. 1A). Logistic modeling estimated a saturation plateau (K) of 5.82×10^7 effective CpG sites, with 95% of this plateau reached at an EE_N of 0.0195. Genome-wide host DNA coverage showed a similar saturating pattern, approaching a theoretical maximum of ~95%, with 95% of this plateau achieved at an EE_N of 0.0159 (Fig. 1B).

Subsampling analyses of five individuals indicated that CpG site detection also saturated with increasing sequencing depth (Fig. 2). Across individuals, estimated plateau values (K) were highly consistent, with bootstrap mean estimates ranging from 5.5×10^7–6.2×10^7 CpG sites (Table S2). The sequencing depth required to reach 95% of the plateau varied among individuals but was consistently low, ranging from 9.8% to 16.1% (7.84 × to 12.88 ×) of total reads. Beyond these depths, additional sequencing resulted in minimal gains in CpG site detection.

Together, these results indicate that WGBS data yield from noninvasively collected samples is primarily limited by host DNA enrichment efficiency rather than sequencing depth, and that the sequencing effort applied here was sufficient to achieve near-saturated CpG site detection.

3.3. Global methylation of fecal DNA

3.3.1. Global methylation pattern

Globally, a total of 64,092,407 unique cytosine methylation sites were observed across the pooled dataset of all 33 samples, accounting for 2.63% of the total nucleotides (2435,140,571 bases in PtaHapG (Lan et al., 2025)). CpG methylation was the predominant type across the entire genome, taking 87.90 ± 1.99% of all methylation events, while CHG and CHH methylation were very rare, taking only 0.67% ± 0.35% and 0.74 ± 0.35%, respectively. Methylation at CH and CHN sites took a total of 10.70 ± 2.40% (Figure S2). Figure S2 illustrated the number of CPG methylation sites across the chromosomes of 33 samples. Notably, the HiC_scaffold_2 chromosome has the highest number, totaling 5501,397 sites. DNA methylation is unevenly distributed across chromosomes (Fig. 3). Higher methylation levels (dark red) are enriched in specific chromosomal regions, whereas other regions display low methylation density (dark blue). Quantitative analysis revealed that DNA methylation site detection was highly uniform and near-saturated across the genome, with recovery proportions ranging from 82.9% to 99.3% per chromosome (Table 1). This high consistency demonstrates that the WGBS data provides a comprehensive and unbiased representation of the species' methylome. The greater breadth of methylation signals observed in larger chromosomes (e.g., HiC_scaffold_1–3) relative to smaller ones (e.g., HiC_scaffold_18–19) primarily reflects their larger genomic spans, confirming that our method achieves consistent coverage regardless of chromosome size.

Methylation rate across chromosomal genome ranged between 0 and 1, with a mean 0.79 ± 0.27. Over 50% of methylated sites exhibited rates greater than 0.9, and more than 75% of which were above 0.97. This signified that the overall state of the fecal DNA in Amur tiger was hypermethylated. The skewness (S_k) of methylation rate distribution ranged from −2.42 to −1.22, kurtosis (B_k) from −0.12–4.94 and medians from 0.89 to 1, suggesting the distribution was significantly left-skewed in relation to normal distribution

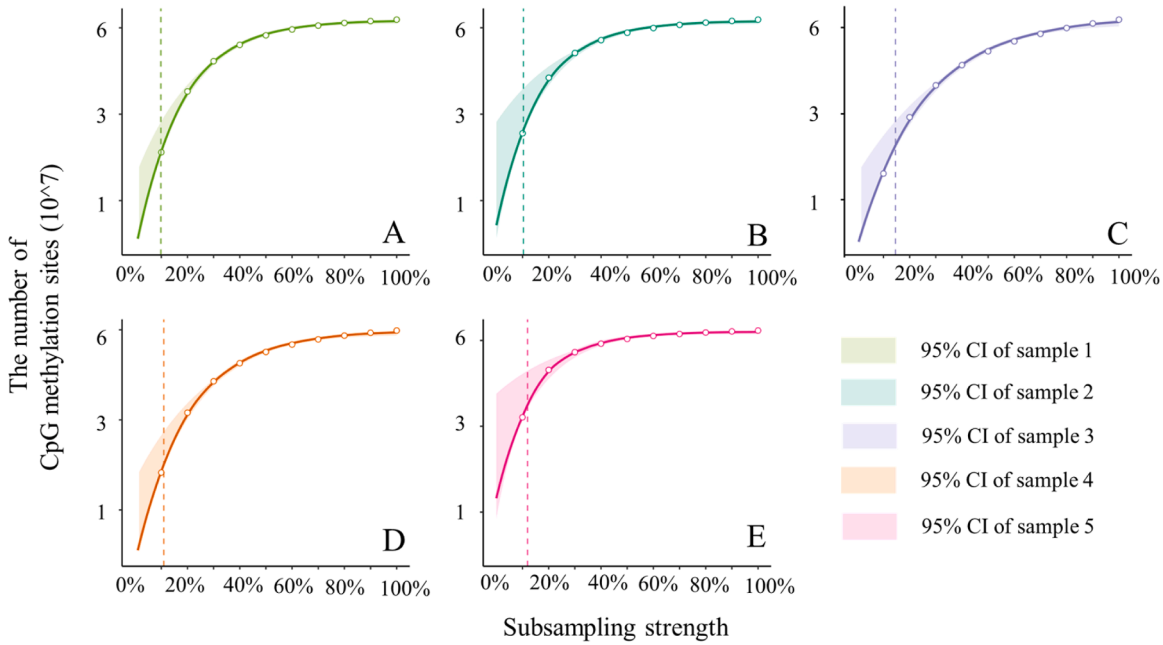

Fig. 2. Subsampling-based evaluation of WGBS sequencing depth. The x-axis indicates the subsampling proportion, and the y-axis shows the corresponding number of CpG methylation sites at each depth. The dashed vertical line indicates the sequencing depth at which the fitted logistic curve reaches 95% of its asymptotic maximum, representing the effective saturation depth for methylation site detection.

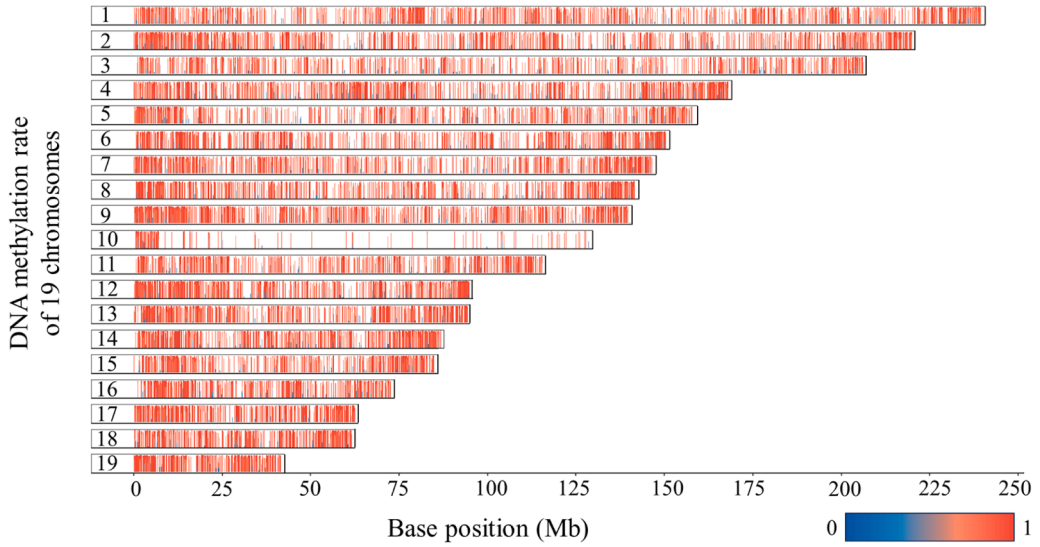

Fig. 3. Chromosomal distribution of DNA methylation rates. The x-axis represents genomic positions (in Mb) along each chromosome, and the y-axis corresponds to DNA methylation rates of different chromosomes arranged from top to bottom. Red vertical bars indicate the presence of methylation sites, with color intensity representing methylation rates. The color gradient from dark blue (low) to red (high) reflects the relative degree of methylation. Numbers 1–19 represent chromosomes.

Table 1
Proportion of recovered CpG methylation sites across 19 chromosomes.

Chromosome ID	Observed Methylation Sites	Total Potential CpG Sites (PtaHapG)	Site Recovery Rate (%)
HiC_scaffold_1	5142757	5241,110	98.12%
HiC_scaffold_2	5501397	5552,222	99.08%
HiC_scaffold_3	4182636	4421,914	94.59%
HiC_scaffold_4	4678303	4723,244	99.05%
HiC_scaffold_5	3345886	3373,910	99.17%
HiC_scaffold_6	3486439	3510,982	99.30%
HiC_scaffold_7	3833542	3891,808	98.50%
HiC_scaffold_8	3667812	3751,486	97.77%
HiC_scaffold_9	3926760	3965,032	99.03%
HiC_scaffold_10	3011224	3630,178	82.95%
HiC_scaffold_11	3088937	3265,430	94.60%
HiC_scaffold_12	3041061	3104,372	97.96%
HiC_scaffold_13	2757783	2844,896	96.94%
HiC_scaffold_14	2723850	2774,020	98.19%
HiC_scaffold_15	2124017	2295,306	92.54%
HiC_scaffold_16	2247653	2581,938	87.05%
HiC_scaffold_17	2627053	2739,308	95.90%
HiC_scaffold_18	2390150	2416,496	98.91%
HiC_scaffold_19	1968924	1999,078	98.49%

(Fig. 4A).

3.3.2. Global methylation heterogeneity

On chromosomal genome, methylation heterogeneity varied from 0.2041 to 0.4593 and averaged 0.3229 ± 0.1584. As shown in the Table S3, the skewness (S_k) of methylation heterogeneity distribution ranged from −0.05–0.90, kurtosis (B_k) from −1.26–0.60 and medians from 0.18 to 0.47, suggesting the distribution was significantly right-skewed in relation to normal distribution mostly (Fig. 4B).

3.3.3. Gene enrichment analysis of age-associated sites

KEGG over-representation analysis of age-associated CpGs showed enrichment in pathways widely linked to ageing and oncogenesis (Fig. 5), with top terms including PI3K–Akt, AGE–RAGE, Rap1, glycosaminoglycan (GAG)-binding proteins, GAG biosynthesis-keratan sulfate (KS), and a dopaminergic module annotated as “cocaine addiction.” Enrichment of PI3K-Akt is consistent with the conserved role of insulin/IGF−1 signaling in lifespan modulation (Kenyon, 2011; López-Otín et al., 2023; Li et al., 2021). Over-representation of AGE–RAGE accords with age-related accumulation of AGEs and downstream inflammatory signaling (Dong

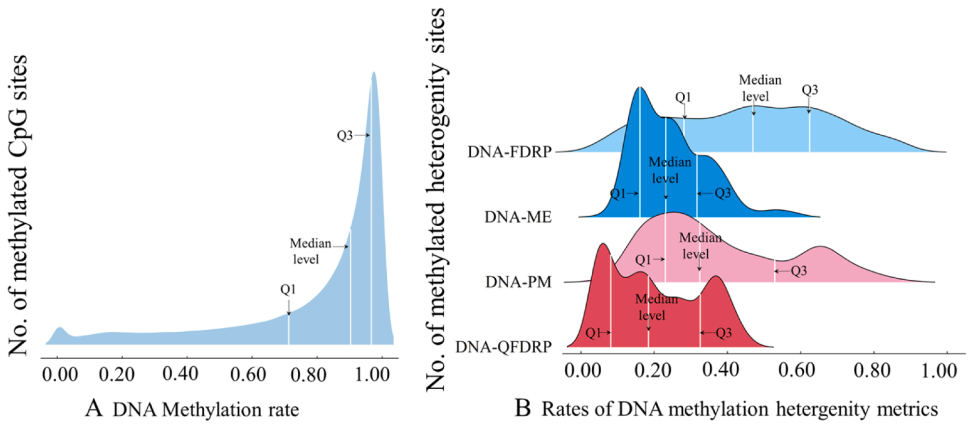

Fig. 4. Distribution patterns of methylation rate and methylation heterogeneity in the nuclear genome. Fig. 3(a): the horizontal axis represented the levels of methylation rate, while the vertical axes displayed the densities of various levels of these methylation rate. Fig. 3(b): the horizontal axis represents the levels of heterogeneity of all samples, and the vertical axes displays the densities of various levels of four methylation heterogeneity metrics.

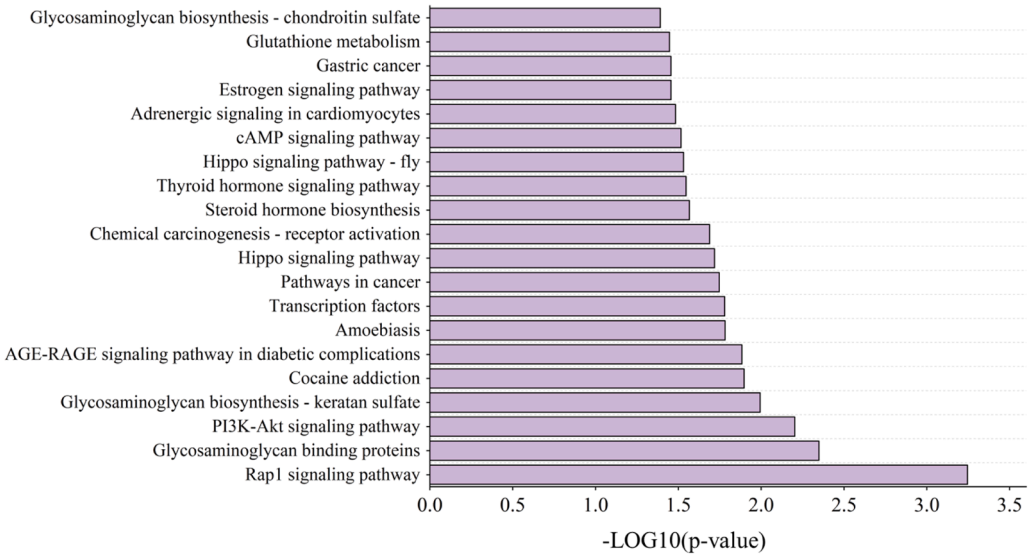

Fig. 5. The pattern of methylation levels of KEGG enrichment of genes associated with age for top 20 functional groups. The x-axis displays -LOG_{10}(p-value), which is calculated by taking the negative logarithm of the original p value, and it serves as a measure of the significance of changes in gene expression within particular pathways, the greater the more significant. The y-axis lists the names of the pathways.

et al., 2022; Lin et al., 2009). Rap1 enrichment highlights endothelial junction/barrier processes that decline with age (Chrzanowska-Wodnicka, 2017; Yamamoto and Watanabe-Takano, 2023). The GAG-related terms (binding proteins and KS biosynthesis) indicate extracellular-matrix remodeling across age (Caterson and Melrose, 2018; Farrugia et al., 2018; Funderburgh, 2000). Finally, the dopaminergic pathway mirrors normative age-related reductions in dopamine system function (Dreher et al., 2008). Collectively, these patterns suggest that age-associated methylation variation concentrates in nutrient-sensing, inflammatory/ECM, vascular-barrier, and neural-*signalling* axes.

3.4. Changes of fecal DNA methylation with age

3.4.1. Spatial genomic location of methylation sites across three age groups

The distribution of age-associated methylation sites across chromosomes exhibited similar patterns among the three age groups (Fig. 6), which was the densest on HiC_scaffold1 and sparest on HiC_scaffold10. Moreover, functional enrichment analysis of the top 20 age-associated loci revealed stage-specific biological implications (Figure S8-S10). In the juvenile group, significantly enriched pathways were predominantly related to developmental and proliferative signaling, including Rap1 signaling, calcium signaling, and Ras/cAMP signaling pathways. In contrast in adult group, the loci were enriched in pathways associated with tissue maintenance and

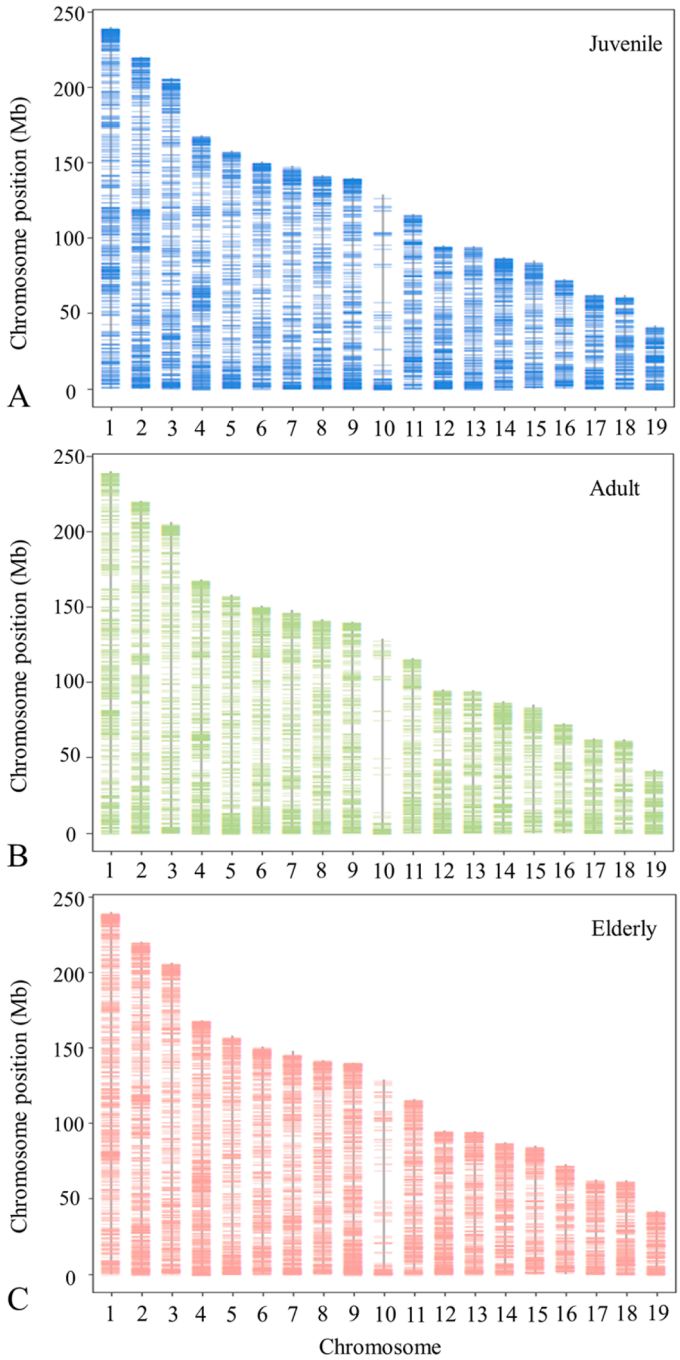

Fig. 6. Chromosomal distribution of age-associated methylation sites in three life stages. Blue bars represent juvenile-specific sites, green bars represent adult-specific sites, and red bars represent elderly-specific sites. Each vertical line indicates the genomic position of an age-associated site. Distinct spatial patterns can be observed across chromosomes, highlighting stage-specific methylation landscapes.

Table 2
Counts of age-associated sites specific to each of three age stages.

Group	Top 200 sites	Top 400 sites	Top 600 sites	Top 800 sites	All sites
Juvenile (n = 5)	187	369	555	745	350,956
Adult (n = 5)	198	386	576	770	374,978
Elderly (n = 5)	192	379	560	747	363,877

cancer-related processes, such as PI3K-Akt signaling, cell adhesion molecules, and pathways in cancer. Elderly-specific loci were mainly enriched in aging- and stress-related pathways, including cellular senescence, cortisol secretion, Cushing syndrome, and the AGE-RAGE signaling pathway in diabetic complications.

A comparison of the top 200, 400, 600, 800, and all age-associated loci further demonstrated that although the absolute numbers of loci increased proportionally across groups, their relative proportions remained low (<0.3%) due to the large background of total methylation sites (Table 2). Interestingly, the expansion from 200 to 800 loci led to increased imbalance among groups, suggesting that the contribution of additional loci may introduce noise rather than enhance classification power.

3.4.2. Difference of methylation rate among age groups

Kruskal-Wallis variance analysis manifested that global chromosomal CpG methylation rate was significantly different among the juvenile, adult and elderly tigers (p = 0.000–0.002) (Figure S3). A total of 13,990, 150,064, and 28,530 differentially methylated regions (DMRs) were identified across three age cohorts namely Juvenile vs. Elderly, Juvenile vs. Adult, and Adult vs. Elderly, respectively (Table S6). The length of 80.0–84.5% DMRs were shorter than 500 bp (see Figure S6), and the genomic distribution patterns were consistent across the cohorts: 79.33–81.15% lying in intergenic regions, 17.06–18.91% in introns, 1.18–1.20% in exons, and 0.56–0.62% in promoters (Figure S5). The distribution of DMRs across chromosomes was heavily skewed, with the top 15 chromosomes containing 86.43–90.77% of all DMRs. Specifically, chromosome HiC_scaffelder_2 had the highest density of DMRs in the Juvenile vs. Elderly (12,096 DMRs) and Juvenile vs. Adult (14,085 DMRs) comparisons, while HiC_scaffelderly_10 had the most DMRs in the Adult vs. Elderly comparison (5080 DMRs) (Figure S7).

3.4.3. Difference of methylation heterogeneity among age groups

As shown in Fig. 7, except for the elderly group, in which the FDRP distribution exhibited a slight left skew (S_K=- 0.0893), all other groups displayed a right-skewed distribution (S_K from 0.0048 to 1.1112). The averages and standard deviations of the four indices progressively shifted leftward with increasing age. In addition to FDRP, the median, Q1, and Q3 values also shifted leftward across age groups (Table S7). Analyses of all four methylation heterogeneity metrics revealed significant differences between the juvenile group and both the adult and elderly groups (P ranges: 4.37E−12–1.45E−02). However, when comparing the adult and elderly groups, a significant difference was only detected in QFDRP (P = 3.03E−03) (Figure S11).

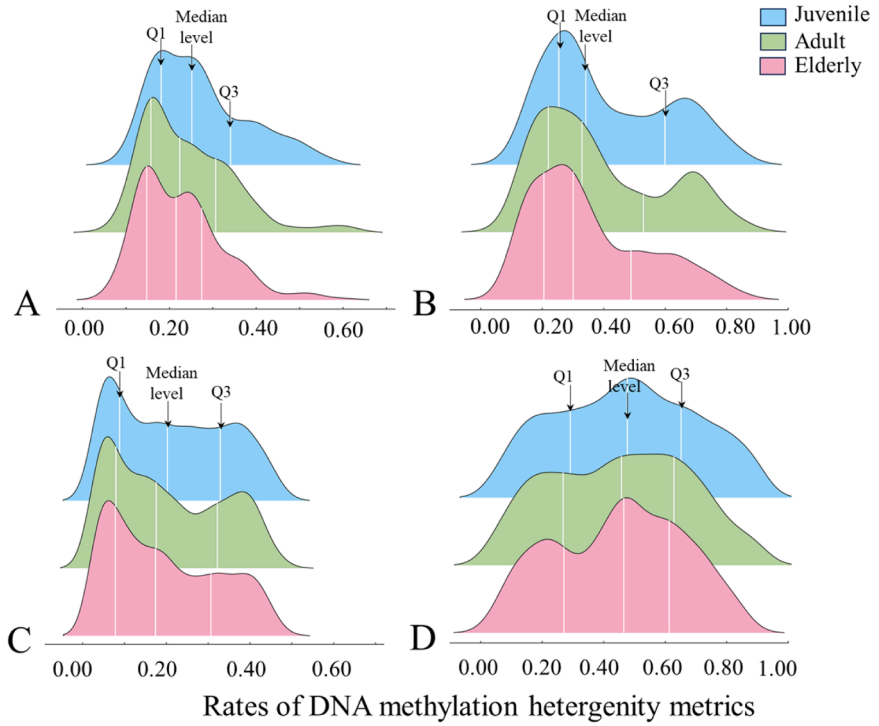

Fig. 7. Distribution patterns of four indices of DNA methylation heterogeneity across the nuclear genome in three age groups. Figure A, B, C, and D illustrats ME, PM, qFDRP, and FDRP, respectively. The horizontal axis represents the levels of DNA methylation heterogeneity. The vertical axes display the frequency of these indices in the three age groups.

3.5. Predicting age using fecal DNA methylation

3.5.1. Clues of age predictability using fecal DNA methylation

Principal component analysis (PCA) revealed that DNA methylation levels, FDRP, and qFDRP could effectively distinguish between the juvenile and elderly groups (Figures S12-S13), although they did not clearly separate the adult group from the juvenile or elderly groups. These results highlight the potential of these indices in predicting age and suggest that combining methylation rate with heterogeneity may more effectively capture age-related information from both dimensions of methylation, thereby improving the robustness of age prediction.

3.5.2. Optimal number of methylation sites and age predictability using methylation rate

3.5.2.1. Feature selection for modeling. We compiled a dataset comprising 472,416 sites obtained from step 5.1, which included DNA methylation levels and chronological ages from 33 fecal samples. After performing Spearman correlation analysis, we identified 5840 sites with correlation coefficients greater than 0.5 or less than −0.5, which were regarded as age-related sites and used to construct a DNA methylation-age model. These methylation sites mainly play four overarching roles including regulation of gene transcription, control of cellular signaling, maintenance of genomic integrity, and modulation of the extracellular environment, according to KEGG enrichment (Fig. 5). Following rigorous screening using Spearman correlation, the quantities of selected qualified CpG sites and the four heterogeneity metric sites, namely ME, PM, FDRP, and qFDRP, were 400, 9, 22, 51, and 30, respectively.

3.5.2.2. Effectiveness of nuDNA methylation rates for age prediction. Using the selected CpG features, three supervised learning models, Elastic Net regression, Support Vector Regression (SVR), and Random Forest (RF), were trained on the 26-sample training set and evaluated in an independent test set (n = 7). All models were tuned using leave-one-out cross-validation with grid search to identify optimal hyperparameters. Across the three algorithms, predicted age showed strong concordance with chronological age in both training and test datasets.

All three models showed strong correspondence between predicted and chronological age. In the training set, Elastic Net, SVR, and RF achieved MAEs of 7.62, 3.89, and 7.70 months, respectively; in the test set, MAEs were 9.15, 7.50, and 10.42 months (Fig. 8). Predicted ages were highly correlated with chronological age across all models (all $R^2 > 0.95$; Fig. 8). SVR consistently exhibited the lowest prediction error in both datasets, identifying it as the best-performing model for age estimation based on the selected CpG sites.

3.5.3. Age predictability using methylation heterogeneity

As summarized in Table S7 and Fig. 9, application of the age-associated CpG sites to the three supervised learning models showed that the random forest regression model achieved the highest predictive accuracy, outperforming both the elastic net and SVM models. In the training set, random forest yielded an R^2 of 0.923, 0.955, 0.962, and 0.961, with a corresponding MAE of ±12.41, ±9.87, ±8.36, and ±9.41 months. In the independent test set, the model maintained strong performance, achieving an R^2 of 0.912, 0.944, 0.971, and 0.966 and an MAE of ±8.34, ±9.33, ±12.34, and ±12.83 months, respectively. When all four age-related metrics were jointly incorporated into the random forest model, prediction accuracy further improved. The training set reached an R^2 of 0.977 with an MAE of ±8.19 months, while the test set achieved an R^2 of 0.867 and an MAE of ±11.10 months.

3.5.4. Predictability of combined model

We attempted to integrate both nuDNA methylation rates and the four heterogeneity metrics into a single predictive model. As shown in Table 3, the R^2 between predicted and actual chronological age ranged from 0.938 to 0.993. The lowest MAE and RMSE were achieved on the training set by elastic regression and support vector machine regression. However, these two indices were substantially higher on the test set, with differences reaching 9.39–10.75 for MAE and 10.62–10.91 for RMSE. In comparison, the random forest regression model demonstrated the balanced performance. It yielded a R^2 of 0.981 and an MAE of ±7.18 months on the training

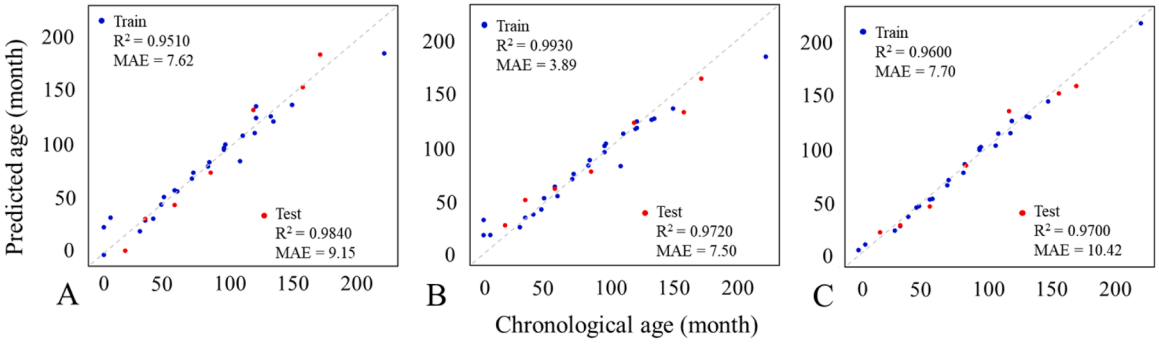

Fig. 8. The relationship between the predicted age and the chronological age of the 400 methylation sites was analyzed using three machine learning models: (A) Elastic regression model, (B) SVM model, and (C) Random forest regression model.

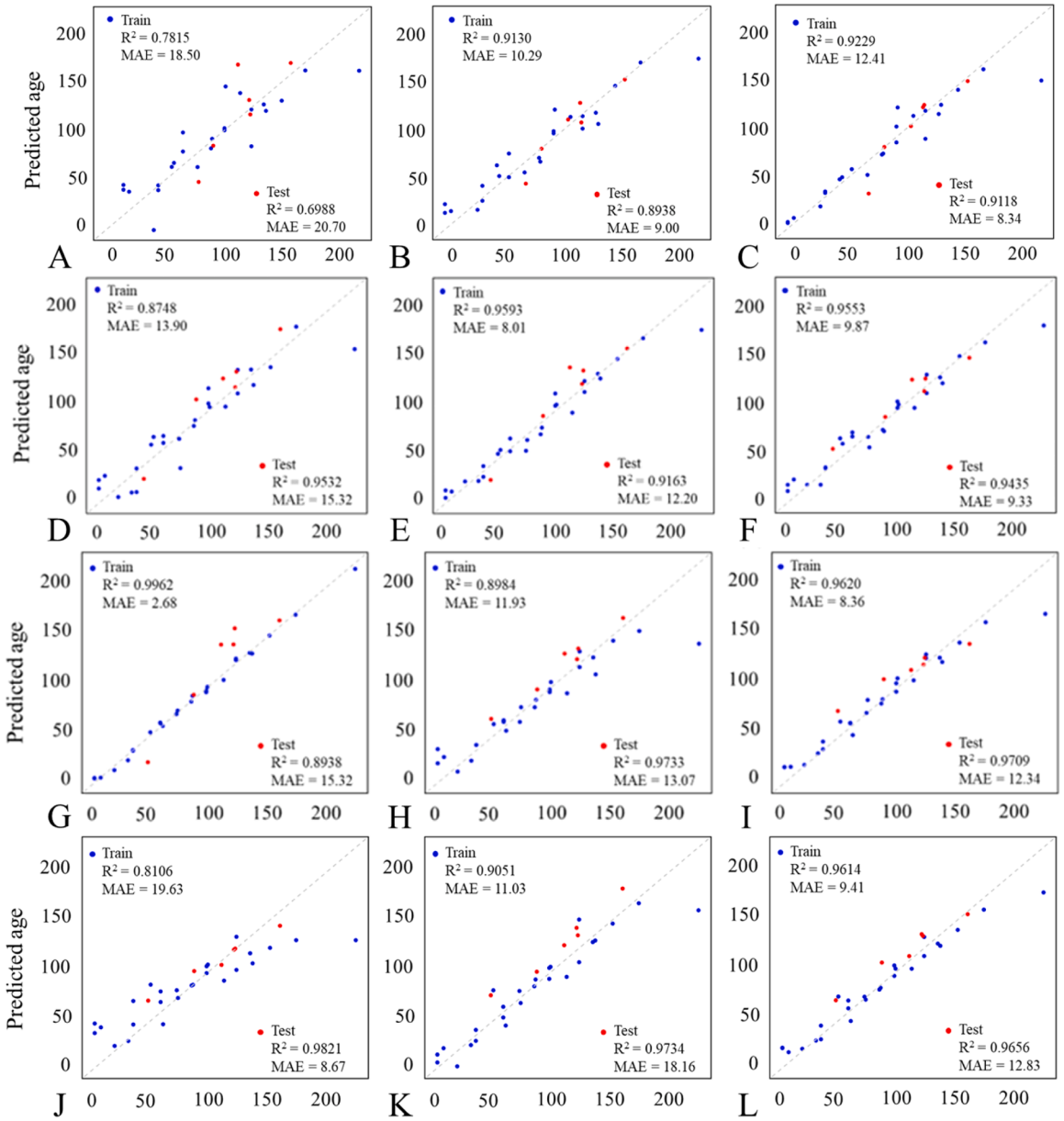

Fig. 9. The correlation between chronological age and predicted age by three machine learning models using ME, PM, FDRP, and qFDRP′ sites. Figure (A-C) show results of the Elastic Net regression, SVM, and Random Forest regression models based on ME sites; Figure (D-F) show corresponding results based on PM sites; Figure (G-I) showd results based on FDRP sites; and Figure (J-L) show results based on qFDRP′ sites.

Table 3
Performance of three machine learning models to predict age by integrating nuDNA methylation rate and heterogeneity metrics.

Models	Data set	MAE/ month	|Dif.|	RMSE/ month	|Dif.|	R^2	|Dif.|
Elastic regression	Train	5.73	10.75	8.27	10.62	0.979	0.041
	Test	16.48		18.89		0.938	
Support vector machine regression	Train	4.24	9.39	4.59	10.91	0.993	0.024
	Test	13.63		15.50		0.969	
Random forest regression	Train	7.18	0.19	10.40	2.17	0.981	0.009
	Test	7.37		8.23		0.972	

Note: MAE is the mean absolute error; RMSE is the root mean squared error; R^2 is the correlation level between predicted and chronological age; |Dif| is the absolute difference between training and test groups.

set, and a R^2 of 0.972 with an MAE of ±7.37 months on the test set. The difference between training and test set was 0.19 in MAE and 2.17 in RMSE. Overall, these results demonstrate that the integrated random forest model outperforms all previous models built using methylation rates or heterogeneity metrics alone.

3.5.5. Effect of sequencing depth on age prediction accuracy

We subsampled the nuDNA data to achieve average sequencing depths of 5 × , 20 × , 40 × , 60 × , and 80 × across the genome. The predictive accuracy (R^2) on the training set remained consistently high, ranging from 0.946 to 1.000, and was not significantly affected by sequencing depth. In contrast, R^2on the test set varied substantially, from 0.122 to 0.984, indicating a strong influence of sequencing depth on generalization performance.

Across all depths, the discrepancy in R^2 between the training and test sets varied widely among the three models. Specifically, at 20 × , the discrepancy ranged from 0.013 to 0.065; it increased to between 0.467 and 0.572 at 40 × , and further to 0.343–0.932 at 60 × . Only at 80 × did all three models achieve high and consistent accuracy, with small discrepancies ranging from 0.009 to 0.033 (Table 5).

Moreover, at 80 × sequencing depth, MAEs were not only the smallest among all tested depths but also showed minimal variation between training and test sets, ranging from 1.53 to 3.61 months. In contrast, MAEs at lower depths varied widely from 6.94 to 59.30 months (Table 4). These results indicated that a sequencing depth of 80 × is necessary to achieve stable and reliable predictive performance.

4. Discussion

4.1. Influence of host DNA enrichment efficiency on CpG recovery and coverage in noninvasive WGBS data

Empirically, environmental exposure (e.g., UV radiation and humidity) of fecal samples accelerates DNA degradation and damage. However, recent methodological advancement to increase the proportion of host nuDNA in the total DNA extract made the sample freshness no longer the absolute determinant of sequencing success (Cui et al., 2024, Yang et al., 2026). This study further confirmed this proportion (herein the host DNA enrichment efficiency, EN_N) is the primary determinant of WGBS data usability. Both the number of effective CpG sites and genome-wide host DNA coverage increased rapidly at low EN_N values and then reached clear saturation as EN_N increased. Logistic modeling estimated an asymptotic plateau of approximately 5.82×10^7 effective CpG sites, with 95% of this plateau reached at an EN_N of 0.0195. Genome-wide host DNA coverage followed a similar pattern, approaching a theoretical maximum of ~95%, with 95% of this plateau achieved at an EN_N of 0.0159. These results indicate that moderate host DNA enrichment, i.e. the ratio of nuclear DNA copy number to bacterial chromosomal DNA copy number to ~1:100, is sufficient to recover the majority of usable epigenomic information from noninvasive samples.

This enrichment-driven saturation has direct implications for epigenetic clock applications. Epigenetic clocks rely on a stable and shared subset of CpG sites across individuals rather than exhaustive genome-wide coverage. The near-consistent upper bound of CpG recovery observed across samples suggests that age-informative CpG sites can be reliably captured once enrichment thresholds are met, providing a sufficient foundation for downstream age estimation. Under these conditions, insufficient host DNA enrichment, rather than sequencing depth, represents the primary risk for unstable CpG representation and reduced cross-sample comparability.

On the other hand, sequencing depth exerts a secondary and diminishing influence on CpG site detection, especially when considerable proportion (even up to 90%) of input DNA is lost during bisulfite treatment (Grunau et al., 2001; Olova et al., 2018). Our subsampling analyses across five individuals revealed highly consistent plateau estimates, with bootstrap mean values ranging from 5.5×10^7 to 6.2×10^7 CpG sites. However, the sequencing depth required to reach 95% of the estimated plateau was low, corresponding to only 9.8–16.1% of total reads. When EN_N exceeds a higher threshold of 0.0195, near-complete CpG recovery and genome-wide coverage can be achieved at an 80 × sequencing depth. This suggests that additional sequencing beyond this point yielded minimal gains in CpG detection. Importantly, once CpG recovery approaches saturation, variation in epigenetic clock performance is unlikely to be driven by the sheer number of detectable CpG sites; instead, factors such as the genomic distribution and functional relevance of CpG sites, methylation variability, and modeling strategies are expected to play a more prominent role. Therefore, utilizing $EN_N \geq 1 \times 10^{-3}$ as an objective baseline quality control filter will provide practical guidance for practice, ensuring the reliability of epigenetic clocks while avoiding aimless sequencing trials on intractable samples.

Lastly, the impact of post-defecation degradation on the methylation status of host DNA is a major concern for accurately capturing

Table 4
Effects of sequencing depth on mean absolute error (MAE) of three models.

Depth	Elastic regression			Support vector machine regression			Random forest regression		
	Train/ month	Test / month	|Dif|	Train/ month	Test / month	|Dif|	Train/ month	Test / month	|Dif|
5 ×	8.91	25.69	16.78	0.10	34.66	34.56	12.34	35.53	23.19
20 ×	6.39	13.33	6.94	3.86	23.28	19.42	10.20	24.77	14.57
40 ×	1.04	58.86	57.82	0.98	60.28	59.3	10.68	59.68	49
60 ×	10.25	48.23	37.98	4.60	46.56	41.96	14.42	41.45	27.03
80 ×	7.62	9.15	1.53	3.89	7.50	3.61	7.70	10.42	2.72

true epigenetic signals in age prediction. However, microbial nucleases in feces may accelerate DNA fragmentation, which primarily reduces data yield rather than altering methylation status. This is because shed epithelial cells in feces tend to lack the cellular machinery required for de novo methylation or demethylation. Moreover, our pipeline's strict alignment to the host reference genome inherently excludes exogenous microbial sequences. Therefore, post-defecation degradation poses a risk of reduced sequencing coverage, rather than inaccuracy in methylation profiling.

4.2. Characteristics of methylation of fecal DNA

Fecal DNA comprises the nuclear and mitochondrial genomes derived from various cell types of the gut epithelium, such as Paneth cells, goblet cells, and enteroendocrine cells (Yang et al., 2025). Each cell type performs specific functions that are influenced or regulated by distinct patterns of genomic methylation (Baba et al., 2022; Grinat et al., 2022; Lin et al., 2025). However, neither the proportion of these cell types within the total exfoliated fecal cells, nor the methylation characteristics of each cell type remains unexplored for wildlife. This gap is primarily due to the lack of efficient methods for isolating and sorting intact cells from feces which is often not fresh enough, and for enriching endogenous DNA, which has limited the application of whole-genome bisulfite sequencing. In this study, we employed our previously developed method, PEERS (Peri-Extraction Enrichment by SDS), to enrich tiger DNA from fecal samples (Cui et al., 2024), and characterized the methylation of the nuclear genome based on bisulfite sequencing of total endogenous DNA. This approach simulates the commonly used approach in research and conservation practice, thus opens new avenues for incorporating fecal epigenetics into ecological research and conservation strategies.

Our methylomic analysis reveals that the genomic landscape of tiger fecal DNA is characterized by a predominantly CpG-context pattern and a state of global hypermethylation, aligning with the typical epigenetic profile of mammalian genomes. The observed global hypermethylation (mean rate 0.79), with the vast majority of sites exhibiting rates exceeding 0.90, reflects the fundamental biological role of DNA methylation in maintaining genomic stability and repressing transposable elements (Smith and Meissner, 2013).

Crucially, our quantitative analysis demonstrates the high technical robustness of using non-invasive fecal samples for WGBS. While absolute counts of methylated sites naturally scaled with chromosome size (e.g., widespread sites on larger scaffolds like HiC_scaffold_1–3 versus fewer on smaller ones like HiC_scaffold_18–19), site recovery was highly uniform and near-saturated across the entire genome (82.9%−99.3% per chromosome; Table 1). The observed fluctuations in methylation density, such as the peak on HiC_scaffold_1 and the minimum on HiC_scaffold_10, therefore likely reflect inherent chromosomal features, such as variations in gene density or chromatin structure, rather than technical biases. This near-saturation confirms that fecal DNA, despite its inherent degradation challenges, provides a comprehensive and unbiased representation of the species' methylome.

Beyond global averages, intra-sample methylation heterogeneity provides critical insights into the epigenetic plasticity and cellular composition of the samples. The right-skewed distribution of heterogeneity (mean 0.32) indicates that while most of the genome remains epigenetically uniform, a distinct subset of loci exhibits high variance (Scherer et al., 2020). In fecal DNA, which comprises a mixture of exfoliated intestinal epithelial cells, this heterogeneity likely captures both the mixed-cell origin and localized epigenetic flexibility responding to environmental or physiological stimuli.

Importantly, this epigenetic plasticity is highly biologically relevant, as demonstrated by the clear separation of methylation traits across age-stratified Amur tiger cohorts. The substantial inter-group dispersions in both global methylation rates and heterogeneity provide quantitative evidence that the tiger methylome undergoes progressive remodeling over time. This dynamic, age-dependent trajectory is consistent with the phenomenon of "epigenetic drift" and the establishment of epigenetic clocks observed in other mammals (Horvath, 2013). The successful capture of these progressive age-related signatures underscores the feasibility of using non-invasive fecal epigenomics to monitor population age structures and physiological states in wild elusive species.

4.3. Changes of methylation with age on fecal DNA

Intestinal epithelial cells, the primary source of host DNA in feces, are characterized by a rapid renewal cycle (Kaunitz and Akiba, 2019). While this high turnover provides a valuable real-time snapshot of transient environmental responses, such as dietary shifts or microbiota fluctuations (Field et al., 2018), capturing long-term aging signatures presents a distinct biological challenge. Our study demonstrates that despite this rapid cellular replacement, stable age-associated epigenetic signals persist in tiger fecal DNA. This persistence is likely driven by systemic epigenetic programming within intestinal stem cells, where cumulative age-related methylation changes are stably inherited and transmitted to rapidly dividing progeny cells (Hannum et al., 2013; Horvath, 2013). Consequently, fecal DNA exhibits a unique dual utility: reflecting short-term physiological plasticity while retaining deeply ingrained, chronological aging marks.

By stratifying the Amur tiger cohort into distinct life history stages (juvenile, adult, and elderly), we successfully isolated a highly specific subset of age-associated CpG sites from background environmental noise. Notably, our genomic annotation revealed that the most robust age-predictive loci were predominantly enriched in non-coding regions, specifically intergenic and intronic sequences, rather than core promoters. This distribution pattern strongly aligns with the mechanisms of "epigenetic drift" and targeted epigenetic regulation. Core promoters are typically strictly protected from hypermethylation to maintain essential housekeeping functions; in contrast, unconstrained non-coding regions, which often harbor enhancers or structural regulatory elements, are more susceptible to accumulating progressive methylation changes over a lifespan (Jones et al., 2015). This suggests that age-related epigenetic rewiring in tigers is primarily driven by distal regulatory elements rather than the direct silencing of gene promoters.

Furthermore, the functional trajectories of these highly informative sites mapped directly onto the physiological realities of the tiger's lifespan. The sequential shift in enriched pathways, from developmental and proliferative signaling in growing juveniles, to

tissue homeostasis in maintaining adults, and ultimately to cellular senescence and stress-response pathways in the elderly (Fig. 5 & S8-S10), confirms that our analytical model captured genuine biological aging rather than statistical artifacts. These functional transitions underscore the physiological relevance of using non-invasive fecal DNA to construct reliable epigenetic clocks, demonstrating its potential to assess not only the chronological age but also the physiological health and life-history stage of elusive wild populations.

4.4. Effectiveness of age prediction using fecal DNA methylation

4.4.1. Informativeness of methylation site, rate and heterogeneity

The development of a robust fecal epigenetic clock relies heavily on the strategic selection of loci that transcend mere statistical association to reflect genuine biological aging (Hannum et al., 2013; Horvath, 2013). Our observation that age-informative sites are enriched in regulatory-sensitive regions with intermediate methylation patterns suggests that these loci function as metastable elements, capturing the gradual, cumulative shift in epigenetic status that persists despite the high regenerative turnover of the intestinal epithelium. The finding that these sites exist in an intermediate state (mean β-value ≈ 0.513) likely reflects a "dynamic equilibrium," where the potential for methylation change is maximized, providing the sensitivity required to track age-related drift across individual intestinal stem cell lineages.

However, interpreting fecal DNA methylation necessitates a nuanced understanding of the low signal-to-noise ratio inherent in the gastrointestinal tract. Unlike stable tissues, the fecal methylome is uniquely vulnerable to environmental stressors and rapid epithelial turnover. The potential impact of nutritional status on stem cell differentiation (Yu et al., 2024), alongside the formation of an "epithelial age mosaic", where localized tissue injuries disrupt the uniform progression of enterocyte maturation and junctional formation (Qin et al., 2025), creates significant biological heterogeneity. This mosaic effect essentially imposes a "noise floor" on chronological age prediction, as the methylation snapshot reflects not only age but also the history of recent cellular repair and environmental interactions.

To address these inherent complexities, we advocate for a transition toward a joint modeling approach. By integrating the rate of methylation change, a proxy for the velocity of biological development and aging (Belsky et al., 2020; 2022), with the entropy of methylation patterns, we move beyond the limitations of static mean methylation. Increased epigenetic entropy, as evidenced in recent literature (Chan et al., 2025; Karetnikov et al., 2024; Seale et al., 2024), serves as a marker of the progressive loss of regulatory fidelity that characterizes cellular senescence. By synthesizing these dimensions, our strategy mirrors the paradigm shift observed in second-generation epigenetic clocks (Lu et al., 2019), which prioritize the capture of biological variability over purely linear correlations. This multidimensional approach is essential not only to compensate for the noise induced by the gut microenvironment but also to extract a more accurate, functionally relevant measure of biological aging from fecal samples, thereby bridging the gap between descriptive epigenetic snapshots and the underlying kinetics of the aging process.

4.4.2. Influence of the number of methylated sites on model accuracy

The number of methylated sites used for modeling is a key determinant of prediction accuracy. In this study, 5840 CpG sites were identified as correlated with tiger age (absolute correlation coefficient > 0.5). However, incorporating additional sites beyond a certain point did not improve age prediction performance. Instead, an optimal subset of 400 top age-associated sites yielded the highest predictive accuracy (MAE = 7.50–10.42 months) across all three modeling approaches. This is consistent with previous research, where using a smaller set of genes or methylation sites resulted in better predictive performance (Peters and Gerber, 2023; Robeck et al., 2021).

A central issue observed was the discrepancy in performance metrics (R^2 and MAE) between the training and test sets during leave-one-out cross-validation, with training performance consistently exceeding that of the test sets (Table S8). We attribute this primarily to the limited sample size. Although the 33 tigers covered a broad and evenly distributed age range, the small number of individuals per age interval provided limited and variable age-related features for the models to learn from.

Under such small-sample conditions, as seen in the Table S8, using too few sites (such as 200) may lead to underfitting, limiting the model's ability to capture true biological signals. Conversely, using too many sites ($\geq$600) introduces excessive feature noise, prompting the model to learn spurious, sample-specific correlations rather than generalizable patterns, a classic case of overfitting. Beyond model-specific factors, a fundamental limitation lies in the moderate informativeness of individual age-associated CpG sites (Table 2), which convey both age-related signals and substantial biological or technical noise. Therefore, using the optimal number of sites and expanding the tiger sample size in the future will be essential to enhance not only prediction accuracy but also, more importantly, the robustness and generalizability of our model.

4.4.3. Influence of sequencing on model accuracy

Fecal samples vary greatly in quality of DNA that often leads to variation of sequencing depth among samples or along the whole genome within a sample (De Flamingh et al., 2022). Sequencing depth at methylated sites is a critical factor in methylation analysis. Low sequencing depth often leads to high uncertainty, increased noise, and a greater risk of misinterpreting technical artifacts as biological signals. In contrast, higher depth improves confidence, reduces noise, and provides a clearer view of true biological heterogeneity (Libertini and Heath, 2016). This presents a trade-off between analytical precision and cost. Estimating the minimum depth required for reliable computation is therefore essential, particularly for fecal DNA, in which a substantial proportion of reads originate from microbial sources, thereby reducing the effective read count available for host epigenetic analysis.

The fecal epigenetic clock we developed for tigers incorporates both methylation rate and heterogeneity metrics. All these metrics

are sensitive to sequencing depth to reveal the true state of a cell population by providing sufficient representation of all genomes in the population, particularly for heterogeneity metrices that measure more complex age-associated features than methylation level alone (Scherer et al., 2020). Using a subsampling approach, we evaluated the effect of a range of average sequencing depths (from 5 × to 80 ×) on regression performance (R^2) and mean absolute error (MAE) between training and test sets. Our results indicate that a sequencing depth of 20 × yields a mean absolute error (MAE) of 13.33 months, providing sufficient resolution for age estimation based on fecal DNA methylation. Increasing the sequencing depth further enhances predictive accuracy, with the MAE reaching its lowest value of 7.50 months at 80 × (Table 4 & 5). We therefore recommend a minimum sequencing depth of 80 × for endogenous DNA to ensure reliable and consistent age prediction.

In this study, feature site selection was performed independently at each sequencing depth, but the number of selected sites was fixed at 400 to maintain comparability across depths. However, the signal-to-noise ratio varies substantially among different depths, and intermediate depths such as 40 × and 60 × may not provide data quality sufficient to support a feature set of this size. For these noisier or less stable depths, using 400 CpG sites likely results in excessive model complexity, enabling the model to fit the training data extremely well while considerably impairing its generalization performance on the test set. This may partially explain the pronounced degradation observed at 40 × and 60 × , where test R^2 dropped markedly despite high training performance. Future work may benefit from depth-specific optimization of feature site number to achieve improved stability and generalization.

Another influential factor is genome-wide coverage, which determines the amount of age-related information that can be retrieved. In our study, the average whole-genome coverage was 83.73% at 80 × depth (sequencing depth). Our models require only a small number, 400 ones here, of selected age-associated sites instead of massive sites to achieve an optimal accuracy. Thus, the genomic coverage was still supportive for chronological age prediction to a precision of 7.50 months (Table S8). This feature well-suited for fecal DNA, which often poses challenges in achieving high-depth whole-genome coverage.

4.4.4. Effectiveness of model

Our study presents a high-resolution WGBS-based epigenetic clock constructed from 512 genomic loci (400 methylation sites and 112 heterogeneity sites), which achieves markedly higher predictive accuracy ($R^2 = 0.972$, MAE = 7.37 months). Previous fecal-based clocks have relied on low-dimensional methylation sites and single-model frameworks. For instance, Hanski et al. (2024) developed a fecal-derived clock for house mice using only 22 CpG sites across five genes and reported prediction errors of approximately 7 days in laboratory mice. Likewise, Yagi et al. constructed an Indo-Pacific bottlenose dolphin clock using only two gene regions measured by methylation-sensitive HRM, with an MAE of 5.08 years, equivalent to approximately 10% of the species' lifespan (Yagi, Qi, 2024). These targeted, low-dimensional designs are constrained by prior knowledge of age-related loci and often fail to capture nonlinear or genome-wide epigenetic dynamics.

Fecal microbiome-based age prediction has been shown to enable only coarse classification of individuals into broad age groups (subadult, adult, and old) by identifying age-associated microbial signatures, as demonstrated in Amur tigers (Hu et al., 2025). While this non-invasive approach provides a useful framework for age grouping, it does not support precise or continuous age estimation and exhibits limited feasibility in wild populations. Fecal microbial communities are strongly influenced by host species, seasonal dietary variation, habitat conditions, and sex (Ma, 2024; Neha and Salazar-Bravo, 2023; Sun et al., 2021), and may further undergo rapid post-defecation alteration under field conditions where timely sample collection cannot be guaranteed (Hale et al., 2016).

In contrast, our WGBS framework leverages comprehensive genome-wide CpG coverage, enabling the detection of subtle and nonlinear age-associated patterns that targeted panels cannot resolve. By integrating Spearman correlation filtering with multiple machine-learning algorithms (ElasticNet, SVM, Random Forest) and combining both methylation level and heterogeneity information, our model shows superior predictive performance, robustness across samples, and enhanced sensitivity to nonlinear methylation trajectories. Importantly, because our data-driven approach does not depend on a priori knowledge of age-related loci, it overcomes the species-specific limitations inherent in targeted approaches. This provides a scalable template for constructing high-resolution clocks in non-model organisms, particularly those lacking pre-established epigenetic arrays or validated CpG marker panels. Even for species lacking highly contiguous reference genomes, this framework can be adapted by mapping reads to closely related surrogate genomes, allowing for the unbiased discovery of conserved aging loci. Collectively, these improvements highlight the power of WGBS-driven, multi-feature, multi-model designs to advance non-invasive age estimation beyond the constraints of earlier low-dimensional methodologies.

5. Future prospects

Estimating wildlife age using non-invasive materials is a longstanding goal for scientists and conservation practitioners. While fecal DNA has been explored for this purpose, its successful application faces two major constraints. The primary bottleneck is the enrichment of endogenous DNA from fecal samples, without which, analysis must rely on amplification-based methods that target only a limited number of age-associated sites resulting in restricted accuracy regardless of their informativeness (Yagi, Qi, 2024). Fortunately, several alternative strategies for host DNA enrichment from non-invasive samples have recently been developed and optimized (Orkin et al., 2021; Tyagi et al., 2024; Sadoughi et al., 2025). In this study, methods such as PEERS (Cui et al., 2024) have proven effective in sequencing the methylome of tiger fecal DNA. However, a major caveat of our current approach is the financial cost of large-scale bisulfite sequencing. At present, the workflow used in this study costs approximately $570.1 per sample (including $0.1 for DNA extraction and PEERS enrichment, and $570 for library preparation and sequencing). While technically feasible and justifiable for highly endangered, well-funded species like tigers, this cost limits immediate large-scale application for wildlife management of less-funded species. To improve upon this, future studies must cross-compare these diverse enrichment methods to maximize yield, and

Table 5
Effects of sequencing depths on age predicting accuracy (R^2) of three models.

Depth	Elastic regression			Support vector machine regression			Random forest regression		
	Train	Test	|Dif|	Train	Test	|Dif|	Train	Test	|Dif|
5 ×	0.982	0.948	0.034	1.000	0.561	0.439	0.968	0.582	0.386
20 ×	0.980	0.915	0.065	0.946	0.931	0.015	0.962	0.949	0.013
40 ×	1.000	0.533	0.467	0.991	0.419	0.572	0.957	0.449	0.508
60 ×	0.989	0.122	0.867	0.996	0.064	0.932	0.971	0.628	0.343
80 ×	0.951	0.984	0.033	0.993	0.973	0.020	0.961	0.970	0.009

transition from exploratory whole-methylome sequencing to the development of targeted, multiplexed assays (e.g., targeted bisulfite sequencing or methylation-specific PCR) focusing exclusively on the most informative sites identified here. This would dramatically reduce the cost per sample and increase throughput.

A second major caveat of the present study lies in identifying age-associated methylation sites and constructing accurate predictive models based exclusively on captive individuals. A fundamental constraint in wildlife research is the scarcity of wild reference samples with known ages, as determining the age of wild animals is often impeded by difficulties in sampling feasibility, invasiveness, and ethical concerns. Consequently, researchers must rely on captive animals with documented ages. In this context, future studies could benefit from longitudinal or repeated-sampling designs in captive individuals of known age. As illustrated by studies on Asian elephants, incorporating repeated samples collected from the same individuals across multiple time points can increase the effective information available for epigenetic clock construction in species for which large independent sample sizes are difficult to obtain (Arai et al., 2023; Arai and Inoue-Murayama, 2026). In such cases, model performance should preferably be evaluated using leave-one-individual-out cross-validation, rather than leaving out individual samples, to reduce the risk of overestimating predictive accuracy due to non-independence among repeated samples. This framework may be particularly useful for large mammals and other wildlife species where sample availability is inherently constrained, and could be incorporated into captive reference systems before broader application to wild populations. The gut epithelium, being highly responsive to dietary and microbial stimuli, may exhibit divergent aging trajectories in different environments. Animals in captivity are characterized by less variable conditions, specialized diets, and prolonged lifespans compared to wild counterparts (Tidière et al., 2016). As a result, models trained on captive individuals may not fully reflect the aging dynamics of wild populations. To improve the ecological validity of these clocks in the future, it is crucial to incrementally integrate fecal samples from wild individuals whose ages are known through long-term field monitoring or other validated biological proxies.

Nonetheless, for many research and conservation purposes, estimating relative rather than absolute age may suffice. When applying the proposed clock to field surveys, future researchers should consider a practical strategy: establish a standardized reference system based on captive populations, while carefully controlling for variables such as living conditions and genetic backgrounds (Kishimoto et al., 2018; Tyshkovskiy et al., 2023). Practitioners must be aware that direct application to wild individuals might initially yield relative age rankings rather than precise chronological ages. With advancing knowledge of how specific methylation sites relate to developmental and aging processes (Kabacik et al., 2022), it is feasible **for future users** to calibrate chronological age with epigenetic biomarkers within such a reference framework (Bernabeu et al., 2023), thereby improving the transferability of models to wild settings.

Finally, when deploying this clock, researchers must consider the inherent biological properties of the sample. Recent advances in DNA methylation research, particularly the development of universal epigenetic clocks across tissues and species (Lu et al., 2023), highlight the promising prospect of a standardized age prediction tool for wildlife studies. While such a universal model would be highly valuable, its construction demands long-term sample accumulation and extensive validation, which are not affordable at present even with international collaboration. Moreover, age-related methylation patterns often exhibit considerable tissue and cell type specificity (Herzog and Redl, 2025; Lin et al., 2009), which may limit the applicability of universal clocks based on highly conserved sites. Therefore, future users are cautioned against applying this fecal-derived clock directly to other biological materials (e.g., blood or hair) without rigorous cross-tissue validation. Furthermore, as fecal samples contain a heterogeneous mix of host shed cells, future applications should account for potential variations in sample degradation and cell-type composition that might influence methylation readouts. In this context, feces represent one of the most accessible non-invasive materials in wildlife research. Focusing on the development of epigenetic clocks specifically optimized for fecal DNA therefore holds greater immediate promise and practical significance for conservation applications.

6. Conclusion

This study presents the first comprehensive landscape of genomic methylation from fecal samples of 33 Amur tigers, generated via Whole-Genome Bisulfite Sequencing (WGBS). We identified significant age-associated methylation changes in intergenic regions, introns, exons, and promoters. By integrating both methylation levels and heterogeneity at 512 selected sites into a random forest model, we successfully predicted chronological age with a mean absolute error (MAE) of 7.37 months. This approach establishes a robust tool for age prediction using captive reference individuals and demonstrates considerable potential for broader applications in wildlife conservation and age-related research across species.

Author Contributions

Wenhui Wang, Yan Chun Xu, Shu Hui Yang initiated and designed the project. Wenhui Wang, Jianing Chu, Xiaoxi Liu, Qi Zhang, Peng Gao and Ao Zou conducted experimental design and completed all experimental operations and analysed data. Wenhui Wang wrote the manuscript. Yan Chun Xu, Shu Hui Yang revised the manuscript. Dan Liu and Haitao Xu provided the samples in this project.

Ethics Statement

This study was approved by the Ethics Committee of Northeast Forestry University (No. 20200512 on May 12, 2020). All procedures were performed in accordance with the ethical guidelines of the 1975 Declaration of Helsinki. Written informed consent was obtained from all participants included in the study.

Funding

This study was supported by the National Natural Science Foundation of China (No. 32170517), Fundamental Research Funds for the Central Universities (2572020DR10) and Project on the Investigation, Supervision and Industry Regulation of Rare and Endangered Species (2024).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors are grateful to the Heilongjiang Siberian Tiger Park for their kind sample donation. Special thanks go to Prof. Wei Zhang, Associate Prof. Su Ying Bai and Bo Li at College of Wildlife and Protected Area, Northeast Forestry University for their technical assistance, support of lab facilities and valuable comments on experiment design and the manuscript.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.gecco.2026.e04284.

Data availability

The filtered deduplicated reads have been deposited to the CNGB (CNP0007549).and machine-learning modeling are openly available on GitHub at: https://github.com/wwh903503949/epigenetic-clock.

References

Arai, K., Inoue-Murayama, M., 2026. Epigenetic signatures of ageing in Asian elephants revealed by reduced representation bisulphite sequencing. Evolut. Appl. 19 (4), e70236. https://doi.org/10.1111/eva.70236.

Arai, K., Qi, H., Inoue-Murayama, M., 2023. Age estimation of captive Asian elephants (Elephas maximus) based on DNA methylation: An exploratory analysis using methylation-sensitive high-resolution melting (MS-HRM). PLoS. One 18 (12), e0294994. https://doi.org/10.1371/journal.pone.0294994.

Armstrong, D.P., Brooks, R.J., 2014. Estimating ages of turtles from growth data. Chelonian Conserv. Biol. 13 (1), 9–15. https://doi.org/10.2744/ccb-1055.1.

Baba, R., et al., 2022. Paneth cell maturation is related to epigenetic modification during neonatal–weaning transition. Histochem. Cell. Biol. 158 (1), 5–13. https://doi.org/10.1007/s00418-022-02110-3.

Belsky, D.W., et al., 2020. Quantification of the pace of biological aging in humans through a blood test, the DunedinPoAm DNA methylation algorithm. eLife 9, e54870. https://doi.org/10.7554/eLife.54870.

Belsky, D.W., et al., 2022. DunedinPACE, a DNA methylation biomarker of the pace of aging. eLife 11, e73420. https://doi.org/10.7554/eLife.73420.

Bernabeu, E., et al., 2023. Refining epigenetic prediction of chronological and biological age. Genome Med. 15 (1), 12. https://doi.org/10.1186/s13073-023-01161-y.

Caterson, B., Melrose, J., 2018. Keratan sulfate, a complex glycosaminoglycan with unique functional capability. Glycobiology 28 (4), 182–206. https://doi.org/10.1093/glycob/cwy003.

Cattet, M., et al., 2018. Can concentrations of steroid hormones in brown bear hair reveal age class? coy001 Conserv. Physiol. 6 (1). https://doi.org/10.1093/conphys/coy001.

Chan, J., Rubbi, L., Pellegrini, M., 2025. DNA methylation entropy is a biomarker for aging. Aging (Albany NY) 17(3) 685698. https://doi.org/10.18632/aging.206220.

Chen, S., et al., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34 (17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

Chrzanowska-Wodnicka, M., 2017. Rap1 in endothelial biology. Curr. Opin. Hematol. 24 (3), 248–255. https://doi.org/10.1097/MOH.0000000000000332.

Cui, L.Y., et al., 2024. A simple and effective method to enrich endogenous DNA from mammalian faeces. Mol. Ecol. Resour. 24 (4), e13939. https://doi.org/10.1111/1755-0998.13939.

Cuomo, M., et al., 2023. Host fecal DNA specific methylation signatures mark gut dysbiosis and inflammation in children affected by autism spectrum disorder. Sci. Rep. 13 (1), 18197. https://doi.org/10.1038/s41598-023-45132-0.

De Flamingh, A., et al., 2022. Combining methods for non-invasive fecal DNA enables whole genome and metagenomic analyses in wildlife biology. Front. Genet. 13, 1021004. https://doi.org/10.3389/fgene.2022.1021004.
Dong, H., et al., 2022. Pathophysiology of RAGE in inflammatory diseases. Front. Immunol. 13, 931473. https://doi.org/10.3389/fimmu.2022.931473.
Dreher, J.C., et al., 2008. Age-related changes in midbrain dopaminergic regulation of the human reward system. Proc. Natl. Acad. Sci. USA 105 (39), 15106–15111. https://doi.org/10.1073/pnas.0802127105.
Duan, R., et al., 2022. Epigenetic clock: A promising biomarker and practical tool in aging. Ageing Res. Rev. 81, 101743. https://doi.org/10.1016/j.arr.2022.101743.
Farrugia, B.L., et al., 2018. The role of heparan sulfate in inflammation, and the development of biomimetics as anti-inflammatory strategies. J. Histochem. Cytochem. 66 (4), 321–336. https://doi.org/10.1369/0022155417740881.
Feng, H., Conneely, K.N., Wu, H., 2014. A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. Nucleic Acids Res. 42 (8), e69. https://doi.org/10.1093/nar/gku154.
Field, A.E., et al., 2018. DNA methylation clocks in aging: categories, causes, and consequences. Mol. Cell. 71 (6), 882–895. https://doi.org/10.1016/j.molcel.2018.08.008.
Funderburgh, J.L., 2000. Keratan sulfate: structure, biosynthesis, and function. Glycobiology 10 (10), 951–958. https://doi.org/10.1093/glycob/10.10.951.
Galkin, F., et al., 2020. Human gut microbiome aging clock based on taxonomic profiling and deep learning. iScience 23 (6), 101199. https://doi.org/10.1016/j.isci.2020.101199.
Grinat, J., Kosel, F., Goveas, N., 2022. Epigenetic modifier balances Mapk and Wnt signalling in differentiation of goblet and Paneth cells. e202101187 Life Sci. Alliance 5 (4). https://doi.org/10.26508/lsa.202101187.
Grunau, C., Clark, S.J., Rosenthal, A., 2001. Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. Nucleic Acids Res. 29 (13) e65-e65.
Hale, V.L., et al., 2016. Effects of field conditions on fecal microbiota. J. Microbiol. Methods 130, 180–188. https://doi.org/10.1016/j.mimet.2016.09.017.
Hannum, G., et al., 2013. Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol. Cell. 49 (2), 359–367. https://doi.org/10.1016/j.molcel.2012.10.016.
Hanski, E., et al., 2024. Epigenetic age estimation of wild mice using faecal samples. Mol. Ecol. 33 (8), e17330. https://doi.org/10.1111/mec.17330.
Herzog, C.M.S., Redl, E., 2025. Functionally enriched epigenetic clocks reveal tissue-specific discordant aging patterns in individuals with cancer. Commun. Med. (Lond.) 5 (1), 98. https://doi.org/10.1038/s43856-025-00739-4.
Holly, A.C., et al., 2013. Towards a gene expression biomarker set for human biological age. Aging Cell. 12 (2), 324–326. https://doi.org/10.1111/acel.12044.
Horvath, S., 2013. DNA methylation age of human tissues and cell types. R115 Genome Biol. 14 (10). https://doi.org/10.1186/gb-2013-14-10-r115.
Hu, X., et al., 2025. Revealing Amur tiger family pedigrees based on age identification using fecal microbiome and kinship analysis. Front. Microbiol. 16, 1666201. https://doi.org/10.3389/fmicb.2025.1666201.
Ira, R., et al., 2024. Understanding aging through the lense of gut microbiome. Explor. Res. Hypothesis Med. 9(4) 294307. https://doi.org/10.14218/erhm.2024.00008.
Jansen, R., et al., 2021. An integrative study of five biological clocks in somatic and mental health. eLife 10, e59479. https://doi.org/10.7554/eLife.59479.
Jones, M.J., Goodman, S.J., Kobor, M.S., 2015. DNA methylation and healthy human aging. Aging Cell. 14 (6), 924–932. https://doi.org/10.1111/acel.12349.
Kabacik, S., et al., 2022. The relationship between epigenetic age and the hallmarks of aging in human cells. Nat. Aging 2 (6), 484–493. https://doi.org/10.1038/s43587-022-00220-0.
Karetnikov, D.I., et al., 2024. Age prediction using DNA methylation heterogeneity metrics. Int. J. Mol. Sci. 25 (9), 4967.
Kaunitz, J.D., Akiba, Y., 2019. Control of intestinal epithelial proliferation and differentiation: the microbiome, enteroendocrine L cells, telocytes, enteric nerves, and GLP, too. Dig. Dis. Sci. 64 (10).
Kenyon, C., 2011. The first long-lived mutants: discovery of the insulin/IGF-1 pathway for ageing. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 366 (1561), 9–16. https://doi.org/10.1098/rstb.2010.0276.
Kishimoto, S., Uno, M., Nishida, E., 2018. Molecular mechanisms regulating lifespan and environmental stress responses. Inflamm. Regen. 38, 22. https://doi.org/10.1186/s41232-018-0080-y.
Krištić, J., et al., 2014. Glycans are a novel biomarker of chronological and biological ages. J. Gerontol. Ser. A Biol. Sci. Med. Sci. 69 (7), 779–789. https://doi.org/10.1093/gerona/glt190.
López-Otín, C., et al., 2023. Hallmarks of aging: An expanding universe. Cell 186 (2), 243–278. https://doi.org/10.1016/j.cell.2022.11.001.
Lan, T., et al., 2025. Revealing extensive inbreeding and less efficient purging of deleterious mutations in wild Amur tigers in China. J. Genet. Genom. 52 (5), 641–649. https://doi.org/10.1016/j.jgg.2024.12.004.
Lee, D., et al., 2023. Metheor: Ultrafast DNA methylation heterogeneity calculation from bisulfite read alignments. PLOS Comput. Biol. 19 (3), e1010946. https://doi.org/10.1371/journal.pcbi.1010946.
Li, W., Bao, H., Zhang, M., 2017. Habitat analysis and design of potential ecological corridors for Amur tiger in Northeastern China. Acta Theriol. Sin. 37 (4), 317–326. https://doi.org/10.1038/s41467-021-24816-z.
Li, W.J., et al., 2021. Insulin signaling regulates longevity through protein phosphorylation in Caenorhabditis elegans, 12 (1), 4568.
Li, Y., et al., 2016. Habitat suitability evaluation for Amur tigers in northeast China based on GIS and RS. J. Zhejiang A&F. Univ. 33 (2), 265–271.
Libertini, E., Heath, S.C., 2016. Saturation analysis for whole-genome bisulfite sequencing data. Nat. Biotechnol. 34 (7), 691–693. https://doi.org/10.1038/nbt.3524.
Lin, L., Park, S., Lakatta, E.G., 2009. RAGE signaling in inflammation and arterial aging. Front. Biosci. (Landmark Ed.) 4 (14), 1403–1413. https://doi.org/10.2741/3315.
Lin, Q., et al., 2025. Colonic epithelial-derived FGF1 drives intestinal stem cell commitment toward goblet cells to suppress inflammatory bowel disease. Nat. Commun. 16 (1), 3264. https://doi.org/10.1038/s41467-025-58644-2.
Lowe, R., et al., 2020. DNA methylation clocks as a predictor for ageing and age estimation in naked mole-rats, Heterocephalus glaber. Aging (Albany NY) 12 (5), 4394–4406. https://doi.org/10.18632/aging.102892.
Lu, A.T., et al., 2019. DNA methylation GrimAge strongly predicts lifespan and healthspan. Aging (Albany NY) 11 (2), 303–327.
Lu, A.T., Fei, Z., Haghani, A., 2023. Universal DNA methylation age across mammalian tissues. Nat. Aging 3 (9), 1144–1166. https://doi.org/10.1038/s43587-023-00462-6.
Luo, S.J., Liu, Y.C., Xu, X., 2019. Tigers of the world: genomics and conservation. Annu. Rev. Anim. Biosci. 7, 521–548. https://doi.org/10.1146/annurev-animal-020518-115106.
Ma, X., 2024. Spatiotemporal differences induced changes in the structure and function of the gut microbiota in an endangered ungulate. Anim. Micro 6 (1), 74. https://doi.org/10.1186/s42523-024-00362-z.
Menni, C., et al., 2015. Circulating proteomic signatures of chronological age. J. Gerontol. Ser. A Biol. Sci. Med. Sci. 70 (7), 809–816. https://doi.org/10.1093/gerona/glu121.
Milner-Gulland, E.J., Akçakaya, H.R., 2001. Sustainability indices for exploited populations. Trends Ecol. & Evol. 16 (12), 686–692.
Moqri, M., Poganik, J.R., 2025. What makes biological age epigenetic clocks tick. Nat. Aging 5 (3), 335–336. https://doi.org/10.1038/s43587-025-00833-1.
Mutz, J., Iniesta, R., 2024. Metabolomic age (MileAge) predicts health and life span: A comparison of multiple machine learning algorithms. eLife 10 (51), eadp3743. https://doi.org/10.1126/sciadv.adp3743.
Neha, S.A., Salazar-Bravo, J., 2023. Fine-scale spatial variation shape fecal microbiome diversity and composition in black-tailed prairie dogs (*Cynomys ludovicianus*). BMC Microbiol. 23 (1), 51. https://doi.org/10.1186/s12866-023-02778-0.
Olova, N., et al., 2018. Comparison of whole-genome bisulfite sequencing library preparation strategies identifies sources of biases affecting DNA methylation data. Genome Biol. 19 (1), 1–22.
Orkin, J.D., et al., 2021. The genomics of ecological flexibility, large brains, and long lives in capuchin monkeys revealed with fecalFACS. Proc. Natl. Acad. Sci. USA 118 (7), e2010632118. https://doi.org/10.1073/pnas.2010632118.

Pacheco, C., Bustamante, C., Araya, M., 2021. Mass-effect: Understanding the relationship between age and otolith weight in fishes. Fish Fish 22 (3), 623–633. https://doi.org/10.1111/faf.12542.
Peters, K.J., Gerber, L., 2023. An epigenetic DNA methylation clock for age estimates in Indo-Pacific bottlenose dolphins (*Tursiops aduncus*). Evolut. Appl. 16 (1), 126–133. https://doi.org/10.1111/eva.13516.
Peters, M.J., et al., 2015. The transcriptional landscape of age in human peripheral blood. Nat. Commun. 6, 8570. https://doi.org/10.1038/ncomms9570.
Qi, H., et al., 2021. Age estimation using methylation-sensitive high-resolution melting (MS-HRM) in both healthy felines and those with chronic kidney disease. Sci. Rep. 11 (1), 19963. https://doi.org/10.1038/s41598-021-99424-4.
Qin, P., et al., 2025. Age mosaic of gut epithelial cells prevents aging. Nat. Commun. 16 (1), 6734. https://doi.org/10.1038/s41467-025-62043-y.
Robeck, T.R., Fei, Z., Lu, A.T., 2021. Multi-species and multi-tissue methylation clocks for age estimation in toothed whales and dolphins, 4 (1), 642. https://doi.org/10.1038/s42003-021-02179-x.
Sadoughi, B., et al., 2025. Non-invasive measures of DNA methylation capture molecular aging in wild capuchin monkeys. bioRxiv. https://doi.org/10.1101/2025.06.03.657357.
Scherer, M., et al., 2020. Quantitative comparison of within-sample heterogeneity scores for DNA methylation data. Nucleic Acids Res. 48 (8), e46. https://doi.org/10.1093/nar/gkaa120.
Seale, K., et al., 2024. A comprehensive map of the aging blood methylome in humans. Genome Biol. 25 (1), 240. https://doi.org/10.1186/s13059-024-03381-w.
Shaochun, Z., et al., 2008. Regional distribution and population size fluctuation of wild Amur tiger (Panthera tigris altaica) in Heilongjiang Province. Acta Theriol. Sin. 28 (2), 165–173.
Shen, W., et al., 2016. SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. PLOS ONE 11 (10), e0163962. https://doi.org/10.1371/journal.pone.0163962.
Smith, Z.D., Meissner, A., 2013. DNA methylation: roles in mammalian development. Nat. Rev. Genet. 14 (3), 204–220. https://doi.org/10.1038/nrg3354.
Stander, P.E., 2015. Field age determination of leopards by tooth wear. Afr. J. Ecol. 35 (2), 156–161.
Sun, G., et al., 2021. Analysis of gut microbiota in three species belonging to different genera (Hemitragus, Pseudois, and Ovis) from the subfamily Caprinae in the absence of environmental variance. Ecol. Evol. 11 (17), 12129–12140. https://doi.org/10.1002/ece3.7976.
Thompson, M.J., et al., 2017. An epigenetic aging clock for dogs and wolves. Aging (Albany NY) 9 (3), 1055–1068. https://doi.org/10.18632/aging.101211.
Tidière, M., et al., 2016. Comparative analyses of longevity and senescence reveal variable survival benefits of living in zoos across mammals. Sci. Rep. 6, 36361. https://doi.org/10.1038/srep36361.
Tyagi, A., et al., 2024. Faecal samples provide genome-wide insights into anthropogenic impacts on two large herbivore species in central India. Mol. Ecol. https://doi.org/10.1111/mec.17461.
Tyshkovskiy, A., Ma, S., Shindyapina, A.V., et al., 2023. Distinct longevity mechanisms across and within species and their association with aging. e2920 Cell 186 (13), 2929–2949. https://doi.org/10.1016/j.cell.2023.05.002.
Wang, Y., et al., 2023. The fine-scale movement pattern of Amur tiger (*Panthera tigris altaica*) responds to winter habitat permeability. Wildl. Lett. 1 (3), 119–130. https://doi.org/10.1002/wll2.12020.
Wehausen, J.D., O'Brien, C.J., McCullough, D.R., 2024. Reliability of tooth cementum rings to age bighorn sheep: a blind test. Calif. Fish. Wildl. J. 110 (4).
Wei, F.W., et al., 1988. Age determination of wild giant pandas. Acta Theriol. Sin. 8, 161–165.
Wen, D., et al., 2022. Conservation potentials and limitations of large carnivores in protected areas: A case study in Northeast China. Conserv. Sci. Pract. 4 (6), e12693. https://doi.org/10.1111/csp2.12693.
Wilkinson, G.S., Adams, D.M., Haghani, A., et al., 2021. DNA methylation predicts age and provides insight into exceptional longevity of bats. Nat. Commun. 12 (1), 1615. https://doi.org/10.1038/s41467-021-21900-2.
Wright, P.G.R., Mathews, F., 2018. Application of a novel molecular method to age free-living wild Bechstein's bats. Mol. Ecol. Resour. 18 (6), 1374–1380. https://doi.org/10.1111/1755-0998.12925.
Xi, Y.X., Li, W., 2009. BSMAP: whole genome bisulfite sequence MAPping program. BMC Bioinforma. 10, 232. https://doi.org/10.1186/1471-2105-10-232.
Xu, Y.C., Li, B., Li, W.S., et al., 2005. Individualization of tiger by using microsatellites. Forensic Sci. Int. 151 (1), 45–51. https://doi.org/10.1016/j.forsciint.2004.07.003.
Yagi, G., Qi, H., 2024. Non-invasive age estimation based on faecal DNA using methylation-sensitive high-resolution melting for Indo-Pacific bottlenose dolphins. Mol. Ecol. Resour. 24 (2), e13906. https://doi.org/10.1111/1755-0998.13906.
Yamamoto, K., Watanabe-Takano, H., 2023. Rap1 small GTPase is essential for maintaining pulmonary endothelial barrier function in mice. FASEB J. 37 (12), e23310. https://doi.org/10.1096/fj.202300830RR.
Yang, J.C., et al., 2026. Assessment of High Throughput Sequencing Quality of Host DNA Enriched From Faeces: A Case From Captive Tiger. Mol. Ecol. Resour. 26 (3), e70120. https://doi.org/10.1111/1755-0998.70120.
Yang, S., Liu, H., Liu, Y., 2025. Advances in intestinal epithelium and gut microbiota interaction. Front. Microbiol. 16, 1499202.
Yu, G., et al., 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS A J. Integr. Biol. 16 (5), 284–287. https://doi.org/10.3389/fmicb.2025.1499202.
Yu, Z., et al., 2024. Nutrient-sensing alteration leads to age-associated distortion of intestinal stem cell differentiating direction. Nat. Commun. 15 (1), 9243. https://doi.org/10.1038/s41467-024-53675-7.