Title: Variable performance of widely used bisulfite sequencing methods and read mapping software for DNA methylation

Authors: Emily V. Kerns, Jesse N. Weber

Corresponding Author: Emily V. Kerns, University of Wisconsin-Madison, Department of Integrative Biology, Kerns3@wisc.edu, https://orcid.org/0009-0002-0538-4632

Jesse N. Weber, University of Wisconsin-Madison, Department of Integrative Biology, Jnweber2@wisc.edu, https://orcid.org/0000-0003-4839-6684

Funding Statement: This work was funded by a startup award from the University of Wisconsin-Madison Department of Integrative Biology (JNW), UW-Madison Department of Integrative Biology Graduate Summer Research Award (EVK), and National Institute of General Medical Sciences 5R35GM142891 (JNW).

Conflict of Interest Disclosure: We have no conflicts of interest.

Ethics Approval Statement: All animals were handled and euthanized in accordance with UW-Madison IACUC protocol L006460-A01.

Abstract:

DNA methylation (DNAm) is the most commonly studied marker in ecological epigenetics, yet the performance of popular library preparation strategies and bioinformatic tools is seldom assessed in genetically variable natural populations. We profiled DNAm using reduced representation bisulfite sequencing (RRBS) and whole genome bisulfite sequencing (WGBS) of technical and biological replicates from threespine stickleback (*Gasterosteus aculeatus*) liver tissue. We then compared how the most commonly used methylation caller (Bismark) performed relative to two alternative pipelines (BWA mem or BWA meth read mappers analyzed with MethylDackel). BWA meth provided 50% and 45% higher mapping efficiency than BWA mem and Bismark, respectively. Despite differences in mapping efficiency, BWA meth and Bismark produced similar methylation profiles, while BWA mem systematically discarded unmethylated cytosines. Depth filters had large impacts on CpG sites recovered across multiple individuals, particularly with WGBS data. Notably, the prevalence of CpG sites with intermediate methylation levels is greatly reduced in RRBS, which may have important consequences for functional interpretations. We conclude by discussing how library construction and bisulfite sequence alignment software can influence the abundance and reliability of data available for downstream analysis. Our analyses suggest that researchers studying genetically variable populations will benefit from deeply sequencing a few initial individuals to identify the amount of genomic coverage necessary for mean methylation estimates to plateau, a value that may differ by species and population. We additionally advocate for paired end sequencing on RRBS libraries to filter SNPs that may bias methylation metrics, which is counter to conventional wisdom.

Key words: DNA methylation, RRBS, WGBS, Bismark, BWA meth, Ecological epigenetics

Introduction:

Epigenetics is an umbrella term for chemical and conformational modifications to DNA and proteins, with DNA methylation (DNAm) being among the most well studied epigenetic marks in ecologically focused studies (Banta & Richards, 2018; Hu & Barrett, 2017; Lea et al., 2017; Rey et al., 2020). As molecular ecologists increasingly explore the role of methylation in natural systems, there is a need for increased attention to how methodological choices impact inferences of between and within individual variation. This is particularly important given that most methylation profiling methods were originally developed and applied in inbred laboratory model organisms.

There are multiple forms of DNAm, but the most abundant type in vertebrates is the addition of a methyl group to cytosine at CpG dinucleotides, often found in CpG-dense regions of the genome (CpG islands) (Jaenisch & Bird, 2003). DNAm has a broad range of genomic effects, including silencing transposable elements (Marin et al., 2020), blocking transcription factors and other proteins from binding to DNA, increasing the likelihood of spontaneous deamination of methylated cytosines converting them to thymine (Ord et al., 2023), and slowing down transcriptional machinery (Cholewa-Waclaw et al., 2019). While its role varies, increased methylation is often associated with decreased gene expression (Zemach et al., 2010). Multigenerational common garden experiments show that although the genomic loci that undergo DNAm are largely heritable, they can also be modified by environmental stimuli (Bogan et al., 2023; Heckwolf et al., 2020; Hu et al., 2021). Moreover, environmentally-induced shifts in DNAm can be inherited across generations (Takahashi et al., 2023; Tobi et al., 2018). Recent findings also suggest that methylated regions experience different selection pressures than background genomic DNA (Ord et al., 2023). Despite accumulating evidence that DNAm may

play an important role in evolution, this topic is still an active area for debate (Adrian-Kalchhauser et al., 2020; Charlesworth et al., 2017; Jaenisch & Bird, 2003; Otterdijk & Michels, 2016). As interest in the functional and evolutionary roles of DNAm continues to increase, more attention needs to be directed toward developing a robust understanding of which loci are methylated and at what rates.

DNAm is commonly measured using bisulfite sequencing methods. Treating DNA with bisulfite converts unmethylated cytosines to uracil, which then becomes thymine during PCR amplification. When compared to a reference genome, C/T mismatches are identified as unmethylated cytosines, while any cytosines remaining are inferred to be methylated. Although DNAm at any particular base pair is a binary trait (i.e., present/absent at a CpG site on each read), the same location may not be methylated across cells or tissues. Some CpG sites will be consistently methylated (or unmethylated) while others will have intermediate methylation (Hay et al., 2023). When bisulfite sequencing is used for bulk sequencing of many cells within a particular tissue, the three possible methylation states of a CpG site in a single cell (present/absent/heterozygous) are summed and become a quantitative metric (i.e., proportion of total sites that are methylated). The ability to accurately capture this quantitative variation is highly contingent on the number of times that a given base pair is represented in the sequencing data, often referred to as read depth.

The two most commonly used bisulfite sequencing methods are whole genome bisulfite sequencing (WGBS) and reduced representation bisulfite sequencing (RRBS; Gu et al., 2011). While WGBS offers the potential to profile all CpG sites, it also requires large amounts of sequencing data in order to repeatedly sample individual DNAm sites. This requirement often limits WGBS to experiments involving few biological replicates and relatively low read depth,

which reduces accuracy of methylation calls and statistical power to detect group-level differences (Seiler Vellame et al., 2021). Additionally, because WGBS does not enrich for methylated sites, much of the sequence data is uninformative for epigenetic studies. For example, 70-80% of mapped reads in human WGBS studies do not contain CpG dinucleotides (Elliott et al., 2015). In contrast, RRBS uses methylation-insensitive restriction enzymes to target sequencing toward CpG islands (e.g., the most commonly used enzyme is MspI, with a cut site of CC/GG). CpG islands are regions of the genome with a high density of CpG sites and believed to be functional hotspots for DNAm (Suzuki & Bird, 2008). Thus, RRBS not only enriches for regions most likely to harbor functionally important methylation differences, but by also profiling less than 10% of the genome for any individual it allows researchers to increase sample sizes and read depth of regions covered (Bock et al., 2010; Laine et al., 2022). The large differences in the types of data recovered from WGBS (wide breadth of the genome sequenced, lower read depth, and smaller sample sizes) and RRBS (specifically targets CpG islands, higher read depth, and larger sample sizes) make each approach better for particular types of research questions. Because evolutionary and ecological studies often assess group-level variation, and therefore require relatively large sample sizes, RRBS is becoming the preferred epigenetic profiling method for these disciplines. However, few RRBS studies have directly compared its results with WGBS, making it difficult to directly assess what biases or limitations are inherent to each method.

Apart from the differences in library construction and experimental design, analysis methods also differ for DNAm data. For example, bioinformatic tools apply different approaches to account for unmethylated cytosines no longer matching the reference genome after bisulfite treatment. Bismark (Krueger & Andrews, 2011), the most common tool for mapping bisulfite

converted sequences (Laine et al., 2022) (Table 1), handles mismatches by producing *in silico* sense (C → T) and antisense (G → A) conversions at all CpG sites before aligning reads to a reference genome that is itself converted using the same process. Reads are considered ambiguous and discarded if they align with multiple versions of the *in silico* converted reference genome. In addition to mapping reads, Bismark also streamlines data analysis by extracting methylation information in a file format (i.e., bedGraph) that is compatible with downstream statistical tools. However, this software relies on Bowtie2 (Langmead & Salzberg, 2012) for read mapping, which often produces lower mapping efficiency (percent of reads mapped to the reference) than other commonly used mapping algorithms such as BWA mem (Harrath et al., 2019). Although tuning Bowtie2 mapping parameters can improve mapping rates, this process can pose a substantial hurdle for novice users. Additionally, generating four intermediate *in silico* conversions for both strands of the reference genome and sample reads leads to longer computational run times and greater memory demands than alternative tools (Farrell et al., 2021; Krueger & Andrews, 2011; Nunn et al., 2021).

BWA meth (Pedersen et al., 2014) is an alternative bisulfite sequence alignment script to Bismark. This software is faster than Bismark because it only performs *in silico* conversion of the reference genome prior to read mapping, not the sampled reads, and uses the BWA mem algorithm (Nunn et al., 2021; Sun et al., 2018). However, a current limitation of BWA meth is that it stops after mapping reads; the tool MethylDackel (https://github.com/dpryan79/MethylDackel) is recommended for extracting methylation information. An advantage of MethylDackel is that it uses overlaps between paired end sequencing data to discriminate between SNPs and unmethylated cytosines. If a site is a bisulfite converted cytosine, the opposite strand should have a G, otherwise it is likely a SNP.

MethylDackel allows the user to specify the maximum proportion of non-Gs permitted on the opposite strand before a site is considered a SNP and excluded from the dataset. Importantly, this could impact estimates of DNAm variation if polymorphic sites are frequently eliminated from analyses. To prevent erroneous filtering of sites based on sequencing errors, the user can also specify the minimum depth the opposite strand must have in order to be used for filtering. As polymorphism data are often unavailable for natural populations, this feature has the potential to increase the reliability of DNAm research on understudied systems.

Here, we provide a comparison of approaches for generating and analyzing DNAm data that are particularly relevant to molecular ecologists. Specifically, we generated WGBS and RRBS data from five populations of threespine stickleback (*Gasterosteus aculeatus*). Four individuals were profiled using both WGBS and RRBS to assess how the methods diverge in: 1) which and how many DNAm sites are recovered, 2) depth of coverage, and 3) distributions of percentage DNAm per site. We then applied several bioinformatic pipelines to characterize the performance of bisulfite alignment software and how they impact the mean and distribution of percent DNAm for each individual. While studies have assessed performance of bioinformatic pipelines for human and plant methylation data (Farrell et al., 2021; Nunn et al., 2021; Sun et al., 2018; Zhou et al., 2024), to our knowledge this has not yet been done in ecologically-focused vertebrate systems. We predicted that RRBS would provide fewer total DNAm sites, greater per-site read depth, and that most DNAm would be found in CpG islands (Bock et al., 2010). We also predicted that BWA mem-based read mappers would perform better than Bowtie-based mappers in terms of total mapping efficiency, and that MethylDackel would efficiently remove polymorphic sites. After explaining our findings, we discuss how to optimize sequencing resources for ecological epigenetics, particularly for systems where DNA polymorphisms are

unknown. Finally, we consolidate information for each of our computational wrappers for bisulfite alignment, providing open access code to encourage researchers to test various read mapping methods, optimize their own bioinformatic pipelines, and increase the availability of data for downstream analysis.

Methods:

Literature review of bisulfite library preparation and analysis methods

To assess how frequently Bismark, BWA meth, MethylDackel, and BWA mem are used to analyze methylation data, we used “software name” AND “DNA methylation” to query Web of Science and Google Scholar databases on August 21, 2024. We consolidated the number of search results from each database in Table 1.

Sample Collection

We sampled wild and lab-reared threespine stickleback from five populations (Table 2). RRBS was applied to all populations, while WGBS was used for common garden fish from two Alaskan populations: Watson and Wik Lakes. Fish were euthanized with an overdose of buffered MS-222 according to IACUC protocol L006460-A01. Field samples were dissected lakeside to avoid introducing transportation stress that may influence DNAm profiles. We dissected and stored liver tissue in RNALater (Sigma-Aldrich). We chose liver as our focal tissue because previous work suggests liver-specific methylation is relatively conserved across vertebrates (Klughammer et al., 2023). Liver also has relatively few cell types (Gumucio & Miller, 1981), which should limit between sample variation due to cell type specific methylation. Two lab fish from each of the Roberts, Gosling, and Sayward populations were exposed to the tapeworm

Schistocephalus solidus as part of a separate study, and one of the Sayward individuals was successfully infected. None of these individuals displayed obvious differences in DNAm patterns, and therefore were retained for further analysis (Supplemental Figure 1).

Sequencing

We extracted DNA from each liver sample using carboxyl coated magnetic beads (BOMB.bio method: Oberacker et al., 2019). Bisulfite conversion, library preparation, and sequencing were performed by Admera Health Biopharm Services. RRBS samples were bisulfite treated and then converted to Illumina sequencing libraries and using Zymo-Seq RRBS Library Kit. WGBS were similarly bisulfite converted using EZ DNA Methylation-Gold Kit (Zymo Research, California, USA) and library preparation was done with xGen™ Methylation-Sequencing DNA Library Preparation Kit (Integrated DNA Technologies, California, USA). Sample quality was assessed using Quibit, agarose gel, and qPCR prior to sequencing. Libraries were sequenced on Illumina Novaseq 6000 using paired end 150bp read chemistry. DNA extractions and sequencing were performed in two batches. Each barcoded RRBS sample was mixed into one of two sequencing pools (Table 2), which produced an average (± SEM) of 10.5 (± 1.3) million and 31.5 (± 15.4) million reads per sample following adapter trimming. All WGBS samples were sequenced in one batch, resulting in 58.5 (± 1.0) million reads per sample after adapter trimming.

Read Mapping and Methylation Calling

We trimmed adapters and barcodes using the -paired, -rrbs (for RRBS samples only, to remove cytosines filled in at the end of reads as an artifact of library preparation), and -q 20 options in Trim Galore! v. 0.6.10 (https://github.com/FelixKrueger/TrimGalore). We then

aligned reads to the threespine stickleback reference genome (v.5; Nath et al., 2021) using default parameters in three programs: BWA mem v. 0.7.17, BWA meth v. 0.2.7, and Bismark v. 0.24.2. Bismark resulted in a low mapping efficiency 28.10 (± 2.26%). We followed the guidance in the documentation and aligned the data in single-end mode with the flag -nondirectional, resulting in a mapping efficiency analogous to other studies. We therefore use the data mapped in single end mode for the rest of the analysis. For BWA mem and BWA meth alignments, samtools v. 1.9 (Danecek et al., 2021) was used to convert SAM outputs to BAM format. Mapping efficiency is included in the Bismark output, and we used Bamtools v.2.5.2 (Barnett et al., 2011) to calculate it for BWA meth and BWA mem runs.

To extract methylation from BWA meth and BWA mem results we used MethylDackel v. 0.5.1 (https://github.com/dpryan79/MethylDackel). Specifically, we filtered SNPs, tabulated per base methylation metrics, and created an output file compatible with methylKit in R (--methylKit) (Akalin et al., 2012). As methylation call accuracy scales logistically with read depth and becomes asymptotic at 10x (Seiler Vellame et al., 2021), we used --minDepth 10 to only retain reads that were sequenced at least 10 times. It can be difficult to differentiate between PCR replicates and true biological duplicates in RRBS data due to biased genomic fragmentation (Laine et al., 2022). Therefore, for the RRBS samples only, we followed recommendations in the MethylDackel documentation and applied the --keepDupes option. Three different SNP filtering cut offs were tested (--maxVaraintFrac 0.25, 0.5, and 0.8) with --minOppositeDepth 5. To prioritize capturing potential methylation variation, we used the most lenient SNP cut off (--maxVariantFrac 0.8) for further analyses.

Statistical analyses

We applied a number of statistical tests in R v. 4.3.2 (R Core Team, 2023) to compare read mapping approaches with both RRBS and WGBS data. We used the nonparametric Friedman test of repeated measures to assess whether the mapping efficiency of individual libraries differed between Bismark, BWA meth, and BWA mem. Effect size was calculated using Kendall's W, in which 0.1-0.3 is considered a small effect, 0.3-0.5 is a moderate effect, and >0.5 is a large effect. Post hoc pairwise comparisons were done using a Wilcox signed-rank test and p-values were corrected for multiple comparisons using the Bonferroni method in rstatix v. 0.7.2 and coin v.1.4-3 (Hothorn et al., 2008; Kassambara, 2023b). Figures were generated using ggpubr v. 0.6.0 (Kassambara, 2023a).

For RRBS samples, which received relatively low or high sequence coverage depending on batch, we used Wilcox rank sum exact tests to test whether batch affected mapping efficiency for each of the three read mapping tools. Results were visualized using ggplot2 v. 3.5.1 (Wickham, 2016) (Supplemental Figure 2). The same procedure was used to analyze the number of mapped reads (Supplemental Figure 3) and average read depth for each read mapping approach.

To determine if the resulting methylation profiles of individuals were in agreement despite differences in read mapping and SNP filtering, we compared both the mean and distribution of percent methylation per individual obtained after using Bismark, BWA meth + MethylDackel, and BWA mem + MethylDackel. We used methylKit v. 1.27.1 (Akalin et al., 2012) to calculate the mean and visualize the distribution of percent methylation for each sample. We then created a linear model in ggpmisc v. 0.6.1 (Aphalo, 2024) to compare the mean percent methylation obtained using each pipeline. If, for every individual, the mean percent methylation obtained from one pipeline reliably matches the mean percent methylation

calculated using another pipeline, then the slope should be 1. After observing a heteroscedastic trend associated with BWA mem, we tested whether variation in the amount of sequencing effort for each sample may explain inconsistencies in mean percent methylation obtained with different bioinformatic pipelines. Specifically, in a post hoc analysis, we visualized the relationship between the residuals of each model, the total number of sequences for each sample, and sequencing batch, predicting that greater sequencing effort would be associated with higher agreement between read mappers, and subsequently residuals close to zero. The association between the absolute value of the residuals and total sequences per sample was tested by comparing the fit of a linear and exponential decay model using AIC scores.

We next assessed differences between RRBS and WGBS. To determine if average read depth per sample varied we used a Wilcox rank sum test. We then used ggvenn v.0.1.10 to tabulate the proportion of CpG sites that were shared (1) between methods within an individual and (2) between individuals using the same method. As we were interested in the breadth of genomic coverage between library preparation methods, we did not filter SNPs prior to this comparison and relaxed the minimum depth to test both 5x and 10x. Finally, the frequency distribution of percent methylation per base was also used to visualize how the overall methylation profile for an individual varied based on the sequencing method.

Results:

Read Mapping: BWA meth outperformed BWA mem and Bismark

BWA meth, BWA mem, and Bismark differed significantly in their mapping efficiency using RRBS data (Friedman test $p = 5.85 \times 10^{-14}$, df = 2, Kendall's W = 0.896). With a mean (± SD) mapping efficiency of 99.2% (± 0.18) , BWA meth outperformed both Bismark (54.4 ±

4.35%) and BWA mem (49.3 ± 4.46%) (Bismark - BWA meth: Wilcox signed-rank test r = 0.872, p = 3.48 x 10^{-10}; BWA mem - BWA meth: Wilcox signed-rank test r = 0.872, p = 3.48 x 10^{-10}). The mapping rate for BWA mem was also significantly lower than Bismark (Bismark - BWA mem: Wilcox signed-rank test r = 0.734, p = 7.5 x 10^{-6}) (Figure 1A). Mapping efficiency for Bismark did not differ between sequencing batches (Wilcox rank sum exact test, W = 142, p-value = 0.732), but there were small but significant batch effects for both BWA meth (Wilcox rank sum exact test, W = 261, p = 2.55 x 10^{-8}) and BWA mem (Wilcox rank sum test, W = 253, p-value = 7.112 x 10^{-7}) (Supplemental Figure 2). Despite the batch effects, BWA meth produced the highest mapping efficiency (>99%, 17.8 ± 13.5 million mapped reads per RRBS sample), so its output was used for downstream analysis. For the WGBS samples, BWA meth again performed the best with a mapping efficiency of 99.6 (± 0.16%), resulting in a mean of 58.6 (± 1.1) million mapped reads. Although we had too few WGBS libraries to make strong statistical inferences, the average WGBS mapping efficiency of BWA mem and Bismark were again lower than BWA meth: 74.6% (± 3.43) and 38.7% (± 1.24), respectively (Figure 1B).

BWA mem artificially inflates methylation

Despite significant differences in mapping efficiency and lack of SNP filtering with Bismark, the mean methylation per putatively methylated site was highly similar between BWA meth and Bismark (slope = 0.935; Figure 2A). BWA mem produced inflated methylation rates compared to the other two methods (BWA meth - BWA mem: slope = 0.119, y-intercept = 89.6, Figure 2C; Bismark - BWA mem: slope = 0.118, y-intercept = 88.9, Figure 2E). This bias was so extreme that the median percent methylation for every sample mapped with BWA mem was 100%. While the relationship between mean percent methylation was roughly linear when comparing BWA meth and Bismark (Figure 2A), there was a heteroscedasticity associated with

sequence coverage when comparing BWA mem to the other read mapping methods (Figure 2C,E). We therefore assessed how variation in sequencing effort for each sample may impact downstream methylation calling. With low sequence coverage there was wide variation between methods. Agreement in methylation calls increases with sequence coverage until approximately 20 million reads per sample, but increasing sequencing effort past this threshold provided little to no improvement (Figure 2B, D, F). An exponential decay model was a better predictor of the data than a linear model for all three methods comparisons (BWA meth-Bismark: ΔAIC = 27.54, BWA meth-BWA mem: ΔAIC = 22.07, Bismark-BWA mem: ΔAIC = 12.19); however, only the BWA meth-Bismark comparison recovered sequencing effort as a significant predictor of reduced residuals (t-value = -3.061, p = 0.00444). Notably, between method variation was substantial even between BWA meth and Bismark, averaging 7.3% but reaching as high as 12.7% difference in mean when samples had less than 20 million sequences (Figure 2B). In both BWA mem comparisons, samples with greater sequencing effort tended to have slightly negative residuals (Figure 2D,F). Meanwhile, samples with greater sequencing effort tend to have slightly positive residuals when comparing Bismark to BWA meth (Figure 2B).

We also examined distribution of percent methylation per base to further characterize how read mapping wrapper influenced which CpG sites were retained for analysis (Figure 3, Supplemental Figures 4-8). BWA mem hardly recovered any CpG sites that were consistently unmethylated (<10% of methylation, Figure 3B) while these sites made up 51.6% (± 14.8) and 43.7% (± 13.9) of the data in BWA meth (Figure 3A) and Bismark (Figure 3C), respectively.

Differences in genomic breadth and depth of coverage between RRBS and WGBS

After applying BWA meth and MethylDackel the average median insert size of RRBS reads was 38.06 (± 2.44) compared to 242.25 (± 22.47) for WGBS. Despite the small insert size,

RRBS resulted in a higher per site depth (mean of 16.39 ± 3.46) than WGBS (12.73 ± 0.06). Batch 1 RRBS samples had a significantly lower read depth than batch 2 (14.73 ± 1.07 versus 19.44 ± 4.22; Wilcoxon rank sum exact test, W = 24.5, p-value = 0.0001151), but also a slightly higher mapping efficiency (99.34 ± 0.00086% versus 99.03 ± 0.0013%; Wilcoxon rank sum exact test, W = 261, p-value = 2.553e-08, Supplemental figure 2). This indicates that the difference in depth of coverage between batches was driven by greater sequencing of batch 2, not by better performance of read mapping software.

We next compared the effects of applying different library preparation methods to technical replicates (i.e. RRBS vs WGBS; Table 3), as well as the similarity of biological replicates from a single population (i.e., RRBS versus RRBS, WGBS versus WGBS; Table 4). Applying a minimum depth filter of 10 reads per site on RRBS and WGBS libraries from the same individual resulted in ≤1.0% shared CpG sites. Lowering the minimum depth to 5 reads caused the proportion of shared CpG sites to increase to 2.25% (± 0.42). Notably, at a minimum of 10x depth, >96% of CpG sites were exclusive to WGBS, highlighting the large difference in breadth of coverage between RRBS and WGBS (Table 3).

When assessing CpG sites shared among individuals using RRBS, the percent of overlap depended on fish population. With a minimum of 10x depth, 13.1% and 27.2% of CpG sites were shared between biological replicates from Watson and Wik Lakes, respectively. This population effect was much reduced with WGBS; 18.8% of CpG sites were shared between Watson Lake individuals and 19.7% between Wik individuals. Lowering the minimum depth to 5x increased the proportion of shared CpG sites using RRBS (28.4% for Watson, 40.1% for Wik) and WGBS (57.6% for Watson, 59.0% for Wik) (Table 4).

Apart from comparing overlap in particular sites, the vast majority of the quantitative CpG sites (i.e. the sites that are not always methylated or unmethylated), are not represented in the RRBS data (RRBS: Figure 3A-C, WGBS: Figure 3D-F). At a minimum depth of 10x, the proportion of CpG sites with methylation rates >10% and < 90% was twice as high in the WGBS samples (55.63 ± 2.49%, Figure 3D) than the RRBS samples (28.56 ± 8.37%, Figure 3A). A higher proportion of intermediately methylated sites were recovered in batch 2 of RRBS sequencing (Supplemental Figures 4-9), but none of these samples were also analyzed via WGBS.

Discussion:

The increasing popularity of RRBS for methylation profiling has coincided with a rise in the use of Bismark as the default method for read mapping and methylation calling, particularly in ecologically focused experiments (Table 1). However, little effort has been dedicated to understanding the strengths and limitations of these methodological choices, particularly in genetically variable, wild organisms. By comparing RRBS and Bismark against less commonly used library preparation and analysis alternatives, we identify several options for researchers to maximize their sequencing resources.

Optimizing bisulfite alignment wrappers for ecological epigenetics

As predicted by previous studies (Farrell et al., 2021; Nunn et al., 2021; Zhou et al., 2024), BWA meth outperformed both BWA mem and Bismark in its ability to map the majority of sequenced reads to a reference genome (Figure 1). However, it is noteworthy that this previous work focused on human and plant data. We found no previous studies assessing bisulfite alignment wrappers in wild-caught/non-model systems. Despite differences in mapping efficiency, BWA meth and Bismark provided consistent estimates of the mean percent of CpG

sites methylated per individual (Figure 2). In contrast, BWA mem substantially inflated estimates of mean methylation rate (Figures 2 & 3). This difference clearly reflects the optimization of particular bioinformatic tools for bisulfite converted data. Because BWA mem does not account for C/T mismatches induced during bisulfite conversion, reads enriched with converted thymines are more likely to differ considerably from the reference sequence and be lost during mapping. Importantly, BWA meth and Bismark only partially address this bias through bioinformatic conversions that improve mapping. Bisulfite conversion damages DNA at unmethylated cytosines during library preparation, which inherently inflates genome-wide methylation levels (Olova et al., 2018). The magnitude of this bias is only recently becoming clear thanks to newer library preparation approaches that forego harsh chemical reactions and instead rely on enzymes to convert unmethylated cytosines (e.g., EM-Seq: Vaisvila et al., 2021). If precise estimates of genomewide methylation are of key concern, then consideration should be given to using these newer, non-bisulfite methods of library preparation. Importantly, EM-seq converts unmethylated cytosines to thymine, making the sequencing data compatible with existing bioinformatic tools for bisulfite converted data. Despite these advances, RRBS remains the most cost effective method and is likely sufficient if the primary goal of a project is to measure or detect large changes in methylation (i.e., those likely to have functional effects or are environmentally-induced). Regardless of the library preparation method chosen, BWA meth or Bismark will likely provide similar results.

For projects focusing on metrics related to per-site methylation, a key goal is to maximize the number of usable sequencing reads. Bismark produced a mapping efficiency only slightly higher than BWA mem, yet it remains the most commonly used read mapping method (Table 1). Continued reliance on Bismark may partly be attributed to it being one of the first bioinformatic

tools designed to handle bisulfite converted data (Table 5). Because Bismark does *in silico* conversions of all reads (sense and antisense, read and reference), its reliability is also agnostic to library directionality. Although Bismark's ability to both map reads and call methylation levels makes it particularly accessible for researchers new to epigenetics, the tool's strict default parameters may cause difficulties when working with paired end reads, as we observed when we initially had a 28% mapping efficiency. Although rarely discussed in the literature, low mapping rates of paired end reads appears to be a relatively common issue with this tool (e.g.,: Nunn et al., 2021; Zhou et al., 2024). Indeed, the developers provide thorough documentation to help users resolve this mapping problem. The guidance highlights failure to trim adapters prior to read mapping, potential contaminants, and high stringency under default parameters, among other possible sources of error. Under default parameters, both paired end reads must map to one unique version of the *in silico* converted reference genome, otherwise the paired end sequence gets discarded. Following the advice in the documentation, and in-line with Nunn et al. (2021) and Zhou et al. (2024), we relaxed this threshold by mapping our reads as non-directional. This resulted in a mapping rate (54%) that is more in-line with previous DNAm studies in stickleback (Heckwolf et al., 2020; Hu et al., 2021).

Despite requiring additional tools for methylation calling, BWA meth resulted in significantly higher mapping rates than Bismark, leading to more data available for downstream analysis. While a previous study found that BWA meth had higher error rates than alternative tools (Farrell et al., 2021), this analysis was based on both directional and nondirectional libraries. It is plausible that the increased error rate was driven by strand bias rather than being a software limitation. Following BWA meth, we used MethylDackel to tabulate methylation and filter SNPs. Conventional wisdom is that RRBS libraries should be single end, as paired end

reads may lead to bias by repeatedly sequencing the same sites (Krueger et al., 2012). However, this bias can be reduced by normalizing the data based on median read depth, and if genetic polymorphisms are unknown, the overlapping nature of paired end reads can be exploited by software like MethylDackel to call SNPs. Additionally, using MethylDackel for SNP filtering reduces errors associated with SNPs falsely being considered unmethylated cytosines. Similar to most studies of ecological model organisms, we did not have genotype data for either the focal individuals or populations, and were thus unable to verify the accuracy of SNP filtering using MethylDackel. In addition to facilitating SNP calling, paired end sequencing ensures high mappability in regions with many genomic repeats, which tend to be enriched in RRBS (Sun et al., 2018).

Recently, there has been an influx of bisulfite alignment wrappers that combine the strengths of BWA meth and Bismark. These software packages use BWA mem for read mapping, are compatible with both directional and nondirectional libraries, and call methylation (Table 5). Although our goal was to assess the most commonly used methods, future efforts should aim to similarly compare the strengths and limitations of new software for ecologically-focused experiments.

RRBS preferentially sequences stable methylation in comparison with WGBS

In addition to assessing how bisulfite alignment wrappers performed with RRBS data, we examined how methylation profiles varied across technical and biological replicates using RRBS and WGBS. Despite receiving relatively low amounts of RRBS data, which drove low read depth, RRBS still reliably resulted in 29% higher sequencing depth than WGBS. However, few CpG sites were commonly sequenced between technical and biological replicates using the two sequencing methods (Tables 3 & 4). In line with previous research (Vaisvila et al., 2021), the

number of overlapping CpG sites between replicates greatly increased when the minimum depth threshold was lower. When comparing the methylation patterns recovered from technical replicates, the most striking difference between RRBS and WGBS was the proportion of intermediately methylated sites recovered with both methods (Figure 3). Intermediate methylation comprised 56% of the CpG sites sequenced with WGBS, while it was only 29% of the CpG sites recovered with RRBS.

One potential explanation for the lack of congruence between sites sequenced with RRBS and WGBS is that our sequencing parameters (i.e., 2x150bp reads) were much longer than the small inserts (~38bp) produced via RRBS. This resulted in only a fraction of the total expected sequence data being useful for analyses. More than doubling the individual sequence effort led to a 36% increase in mean depth per site (i.e., 19x mean depth in batch 2 compared to 14x for batch 1). Importantly, increasing sequence effort also led to greater agreement in downstream methylation calling, regardless of read mapping and methylation calling methods. However, this benefit plateaus at ~20 million reads per sample (Figure 2). A previous study used simulations based cortex DNAm of a mouse model to determine that the relationship between depth of coverage and methylation call accuracy increases rapidly until 10x depth of coverage, at which point the trend becomes asymptotic (Seiler Vellame et al., 2021). Comparing our stickleback results with those of Seiler Vellame et al. (2021) suggests that the optimal target coverage depth may vary among species, and also likely depends on what level of precision is important to investigators.

Perhaps the most striking outcome from using technical replicates to contrast sequencing methods was that the proportion of intermediately methylated sites sequenced with WGBS was almost twice as high as intermediate methylation captured with RRBS (56% versus 29%,

respectively). Methylation can degrade if samples are collected multiple hours after mortality (Rhein et al., 2015) or frozen for a long period of time (Lee et al., 2023), but the liver tissue we sequenced was stored in RNALater immediately upon euthanasia and kept at -70□ for less than one year. Therefore, degradation of intermediately methylated sites is unlikely the main driver of this observation. Rather, there are technical and evolutionary explanations that likely interacted to create this pattern. As previously mentioned, batch 2 RRBS samples recovered a much higher proportion of intermediately methylated sites than batch 1 (Supplemental Figures 4-9), likely driven by more sequencing. However, this higher rate (37.83 ± 5.78%) still appears to be substantially lower than the average of WGBS. This pattern suggests that there may be fundamental, biological differences in the methylation patterns surveyed by each method. Future studies could use technical replicates to confirm this reduction of intermediately methylated sites in RRBS data.

In considering how much sequencing to invest in RRBS, there may be evolutionary arguments not to prioritize capturing intermediately methylated sites. A recent study in mouse cell lines found that loci with intermediate methylation probabilistically, but not reliably, transmitted methyl marks during mitosis. Furthermore, their methylation status had little effect on gene expression (Hay et al., 2023). Another study in wild stickleback found that regions of the genome containing stable methylation (<15% and >85% methylation) had stronger signatures of selection than those containing intermediate methylation (Ord et al., 2023). These studies suggest that because CpG islands are often enriched for methylation sites that have functional importance (Jaenisch & Bird, 2003), they are more likely to be subject to selection, and thus may have increased investment by regulatory machinery to maintain methylation state across all cells. Thus, the lower proportion of intermediately methylated cytosines we sequenced with RRBS,

which targets CpG islands, may be picking up a biologically interesting pattern. Additionally, given that molecular ecologists are likely interested in identifying functionally important epigenetic effects, RRBS may provide a better option for detecting and measuring these changes. As the field of ecological epigenetics continues to expand, we encourage researchers to continue to interrogate which CpG sites are informative for evolutionary processes and consider how the sequencing and bioinformatic methods they use may influence resolution of methylation profiles.

Conclusions:

We take a critical approach to analyzing RRBS and WGBS data using several conventional methods. We highlight the strengths and weaknesses of both molecular and bioinformatic methods, all of which should be considered during experimental design. If SNP filtering in the absence of genotype data is a priority, researchers may consider choosing paired end sequencing, contrary to traditional advice regarding RRBS library construction. Given that RRBS can result in exceedingly small insert sizes and there may be random variation in the amount of sequencing effort for each sample, we suggest using relatively short read lengths that are still within the constraints of mapping software (≥70bp) (Harrath et al., 2019) and deeply sequencing a small number of initial samples to determine the sequencing threshold required to obtain accurate per site methylation metrics. In this study using threespine stickleback, 20 million reads/sample appears to be the optimal trade off between cost and accuracy. Additionally, the most commonly used tool for analyzing bisulfite sequencing data (Bismark) performs significantly worse than alternative methods, such as BWA meth, in terms of percentage of reads mapped. Researchers can increase the availability and reliability of data for downstream analysis by considering alternative analysis pipelines (Table 5). Finally, we

highlight how different library preparation methods recover distinct signatures of per-site methylation, and why these patterns warrant additional investigation for understanding how epigenetics influence evolutionary processes.

Acknowledgements:

We thank Sean Schoville for computing resources, and Kristofer Sasser and Cole Wolf for collecting fish from Vancouver Island. We recognize that the fish populations analyzed here originate from lakes in the unceded territory of the Dena'ina and Coast Salish people and thank them for their continued land stewardship. This work was funded by the Department of Integrative Biology at UW-Madison and National Institute of General Medical Sciences 5R35GM142891-02.

References:

Adrian-Kalchhauser, I., Sultan, S. E., Shama, L. N. S., Spence-Jones, H., Tiso, S., Keller Valsecchi, C. I., & Weissing, F. J. (2020). Understanding "Non-genetic" Inheritance: Insights from Molecular-Evolutionary Crosstalk. *Trends in Ecology and Evolution*, *35*(12), 1078–1089. https://doi.org/10.1016/j.tree.2020.08.011

Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F. E., Figueroa, M. E., Melnick, A., & Mason, C. E. (2012). methylKit: A comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biology*, *13*(10), R87. https://doi.org/10.1186/gb-2012-13-10-r87

Aphalo P (2024). ggpmisc: Miscellaneous Extensions to 'ggplot2'. R package version 0.6.1, https://CRAN.R-project.org/package=ggpmisc.

Banta, J.A., Richards, C.L. (2018). Quantitative epigenetics and evolution. *Heredity*, *121*,

210-224. https://doi.org/10.1038/s41437-018-0114-x

Barnett, D. W., Garrison, E. K., Quinlan, A. R., Strömberg, M. P., & Marth, G. T. (2011). BamTools: A C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics*, *27*(12), 1691–1692. https://doi.org/10.1093/bioinformatics/btr174

Bock, C., Tomazou, E. M., Brinkman, A. B., Müller, F., Simmer, F., Gu, H., Jäger, N., Gnirke, A., Stunnenberg, H. G., & Meissner, A. (2010). Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nature Biotechnology*, *28*(10), 1106–1114. https://doi.org/10.1038/nbt.1681

Bogan, S. N., Strader, M. E., & Hofmann, G. E. (2023). Associations between DNA methylation and gene regulation depend on chromatin accessibility during transgenerational plasticity. *BMC Biology*, *21*(1), 149. https://doi.org/10.1186/s12915-023-01645-8

Charlesworth, D., Barton, N. H., & Charlesworth, B. (2017). The sources of adaptive variation. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1855), 20162864. https://doi.org/10.1098/rspb.2016.2864

Cholewa-Waclaw, J., Shah, R., Webb, S., Chhatbar, K., Ramsahoye, B., Pusch, O., Yu, M., Greulich, P., Waclaw, B., & Bird, A. P. (2019). Quantitative modelling predicts the impact of DNA methylation on RNA polymerase II traffic. *Proceedings of the National Academy of Sciences*, *116*(30), 14995–15000. https://doi.org/10.1073/pnas.1903549116

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008. https://doi.org/10.1093/gigascience/giab008

Elliott, G., Hong, C., Xing, X., Zhou, X., Li, D., Coarfa, C., Bell, R. J. A., Maire, C. L., Ligon, K. L., Sigaroudinia, M., Gascard, P., Tlsty, T. D., Harris, R. A., Schalkwyk, L. C.,

Bilenky, M., Mill, J., Farnham, P. J., Kellis, M., Marra, M. A., Milosavljevic, A., Hirst, M., Stormo, G.D., Wang, T., & Costello, J. F. (2015). Intermediate DNA methylation is a conserved signature of genome regulation. *Nature Communications*, *6*(1), 6363. https://doi.org/10.1038/ncomms7363

Farrell, C., Thompson, M., Tosevska, A., Oyetunde, A., & Pellegrini, M. (2021). BiSulfite Bolt: A bisulfite sequencing analysis platform. *GigaScience*, *10*(5). https://doi.org/10.1093/gigascience/giab033

Gu, H., Smith, Z. D., Bock, C., Boyle, P., Gnirke, A., & Meissner, A. (2011). Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nature Protocols*, *6*(4), 468–481. https://doi.org/10.1038/nprot.2010.190

Gumucio, J. J., & Miller, D. L. (1981). Functional implications of liver cell heterogeneity. *Gastroenterology*, *80*(2), 393–403. https://doi.org/10.1016/0016-5085(81)90732-0

Guo, W., Fiziev, P., Yan, W., Cokus, S., Sun, X., Zhang, M. Q., Chen, P.-Y., & Pellegrini, M. (2013). BS-Seeker2: A versatile aligning pipeline for bisulfite sequencing data. *BMC Genomics*, *14*(1), 774. https://doi.org/10.1186/1471-2164-14-774

Harrath, Y., Mahjoub, A., AbuBakr, F., & Azhar, M. (2019). Comparative Evaluation of Short Read Alignment Tools for next Generation DNA Sequencing. *2019 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, 1–6. https://doi.org/10.1109/3ICT.2019.8910272

Hay, A. D., Kessler, N. J., Gebert, D., Takahashi, N., Tavares, H., Teixeira, F. K., & Ferguson-Smith, A. C. (2023). Epigenetic inheritance is unfaithful at intermediately methylated CpG sites. *Nature Communications*, *14*(1), 5336. https://doi.org/10.1038/s41467-023-40845-2

Heckwolf, M. J., Meyer, B. S., Häsler, R., Höppner, M. P., Eizaguirre, C., & Reusch, T. B. H. (2020). Two different epigenetic information channels in wild three-spined sticklebacks are involved in salinity adaptation. *Science Advances*, *6*(12), eaaz1138. https://doi.org/10.1126/sciadv.aaz1138

Hothorn T., Hornik K., van de Wiel M.A., Zeileis A. (2008). Implementing a class of permutation tests: The coin package. *Journal of Statistical Software, 28*(8), 1-23. https://doi.org/10.18637/jss.v028.i08

Hu, J., & Barrett, R. D. H. (2017). Epigenetics in natural animal populations. *Journal of Evolutionary Biology*, *30*(9), 1612–1632. https://doi.org/10.1111/jeb.13130

Hu, J., Wuitchik, S. J. S., Barry, T. N., Jamniczky, H. A., Rogers, S. M., & Barrett, R. D. H. (2021). Heritability of DNA methylation in threespine stickleback (*Gasterosteus aculeatus*). *Genetics*, *217*(1). https://doi.org/10.1093/GENETICS/IYAB001

Jaenisch, R., & Bird, A. (2003). Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. *Nature Genetics*, *33*(3S), 245–254. https://doi.org/10.1038/ng1089

Kassambara A. (2023a). ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.6.0, https://CRAN.R-project.org/package=ggpubr.

Kassambara A. (2023b). rstatix: Pipe-Friendly Framework for Basic Statistical Tests. R package version 0.7.2, https://CRAN.R-project.org/package=rstatix.

Klughammer, J., Romanovskaia, D., Nemc, A., Posautz, A., Seid, C. A., Schuster, L. C., Keinath, M. C., Lugo Ramos, J. S., Kosack, L., Evankow, A., Printz, D., Kirchberger, S., Ergüner, B., Datlinger, P., Fortelny, N., Schmidl, C., Farlik, M., Skjærven, K., Bergthaler, A., Liedvogel, M., Thaller, D., Burger, P.A., Hermann, M., Distel, M., Distel,

D.L., Kübber-Heiss, A., & Bock, C. (2023). Comparative analysis of genome-scale, base-resolution DNA methylation profiles across 580 animal species. *Nature Communications*, *14*(1), 232. https://doi.org/10.1038/s41467-022-34828-y

Krueger, F., & Andrews, S. R. (2011). Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, *27*(11), 1571–1572. https://doi.org/10.1093/bioinformatics/btr167

Krueger, F., Kreck, B., Franke, A., & Andrews, S. R. (2012). DNA methylome analysis using short bisulfite sequencing data. *Nature Methods*, *9*(2), 145–151. https://doi.org/10.1038/nmeth.1828

Laine, V. N., Sepers, B., Lindner, M., Gawehns, F., Ruuskanen, S., & van Oers, K. (2022). An ecologist's guide for studying DNA methylation variation in wild vertebrates. *Molecular Ecology Resources*, 1755-0998.13624. https://doi.org/10.1111/1755-0998.13624

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Lea, A.J., Vilgalys, T.P., Durst, P.A.P., & Tung, J. (2017). Maximizing ecological and evolutionary insight in bisulfite sequencing data sets. *Nature Ecology and Evolution*, *1*, 1074–1083. https://doi.org/10.1038/s41559-017-0229-0

Lee, N. Y., Hum, M., Tan, G. P., Seah, A. C., Kin, P. T., Tan, N. C., Law, H.-Y., & Lee, A. S. G. (2023). Degradation of methylation signals in cryopreserved DNA. *Clinical Epigenetics*, *15*(1), 147. https://doi.org/10.1186/s13148-023-01565-y

Marin, P., Genitoni, J., Barloy, D., Maury, S., Gibert, P., Ghalambor, C. K., & Vieira, C. (2020). Biological invasion: The influence of the hidden side of the (epi)genome. *Functional Ecology*, *34*(2), 385–400. https://doi.org/10.1111/1365-2435.13317

Nath, S., Shaw, D. E., & White, M. A. (2021). Improved contiguity of the threespine stickleback genome using long-read sequencing. *G3: Genes, Genomes, Genetics*, *11*(2). https://doi.org/10.1093/g3journal/jkab007

Nunn, A., Otto, C., Stadler, P. F., & Langenberger, D. (2021). Comprehensive benchmarking of software for mapping whole genome bisulfite data: From read alignment to DNA methylation analysis. *Briefings in Bioinformatics*, *22*(5), bbab021. https://doi.org/10.1093/bib/bbab021

Oberacker, P., Stepper, P., Bond, D. M., Höhn, S., Focken, J., Meyer, V., Schelle, L., Sugrue, V. J., Jeunen, G. J., Moser, T., Hore, S. R., von Meyenn, F., Hipp, K., Hore, T. A., & Jurkowski, T. P. (2019). Bio-On-Magnetic-Beads (BOMB): Open platform for high-throughput nucleic acid extraction and manipulation. *PLoS Biology*, *17*(1). https://doi.org/10.1371/journal.pbio.3000107

Olova, N., Krueger, F., Andrews, S., Oxley, D., Berrens, R. V., Branco, M. R., & Reik, W. (2018). Comparison of whole-genome bisulfite sequencing library preparation strategies identifies sources of biases affecting DNA methylation data. *Genome Biology*, *19*(1). https://doi.org/10.1186/s13059-018-1408-2

Ord, J., Gossmann, T. I., & Adrian-Kalchhauser, I. (2023). High Nucleotide Diversity Accompanies Differential DNA Methylation in Naturally Diverging Populations. *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msad068

Otterdijk, S. D., & Michels, K. B. (2016). Transgenerational epigenetic inheritance in mammals: How good is the evidence? *The FASEB Journal*, *30*(7), 2457–2465. https://doi.org/10.1096/fj.201500083

Pedersen, B. S., Eyring, K., De, S., Yang, I. V., & Schwartz, D. A. (2014). *Fast and accurate*

alignment of long bisulfite-seq reads (No. arXiv:1401.1129). arXiv. http://arxiv.org/abs/1401.1129

R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rey, O., Eizaguirre, C., Angers, B., Baltazar□Soares, M., Sagonas, K., Prunier, J. G., & Blanchet, S. (2020). Linking epigenetics and biological conservation: Towards a *conservation epigenetics* perspective. *Functional Ecology*, *34*(2), 414–427. https://doi.org/10.1111/1365-2435.13429

Rhein, M., Hagemeier, L., Klintschar, M., Muschler, M., Bleich, S., & Frieling, H. (2015). DNA methylation results depend on DNA integrity—Role of post mortem interval. *Frontiers in Genetics*, *6*. https://doi.org/10.3389/fgene.2015.00182

Seiler Vellame, D., Castanho, I., Dahir, A., Mill, J., & Hannon, E. (2021). Characterizing the properties of bisulfite sequencing data: Maximizing power and sensitivity to identify between-group differences in DNA methylation. *BMC Genomics*, *22*(1), 446. https://doi.org/10.1186/s12864-021-07721-z

Sun, X., Han, Y., Zhou, L., Chen, E., Lu, B., Liu, Y., Pan, X., Cowley, A.W., Liang, M., Wu, Q., Lu, Y., Liu, P., & Berger, B. (2018). A comprehensive evaluation of alignment software for reduced representation bisulfite sequencing data. *Bioinformatics*, *34*(16), 2715-2723. https://doi.org/10.1093/bioinformatics/bty174

Suzuki, M. M., & Bird, A. (2008). DNA methylation landscapes: Provocative insights from epigenomics. *Nature Reviews Genetics*, *9*(6), 465–476. https://doi.org/10.1038/nrg2341

Takahashi, Y., Morales Valencia, M., Yu, Y., Ouchi, Y., Takahashi, K., Shokhirev, M. N., Lande, K., Williams, A. E., Fresia, C., Kurita, M., Hishida, T., Shojima, K., Hatanaka, F.,

Nuñez-Delicado, E., Esteban, C. R., & Izpisua Belmonte, J. C. (2023). Transgenerational inheritance of acquired epigenetic signatures at CpG islands in mice. *Cell*, *186*(4), 715-731.e19. https://doi.org/10.1016/j.cell.2022.12.047

Tobi, E. W., Slieker, R. C., Luijk, R., Dekkers, K. F., Stein, A. D., Xu, K. M., Biobank-based Integrative Omics Studies Consortium, Slagboom, P. E., van Zwet, E. W., Lumey, L. H., & Heijmans, B. T. (2018). DNA methylation as a mediator of the association between prenatal adversity and risk factors for metabolic disease in adulthood. *Science Advances*, *4*(1), eaao4364. https://doi.org/10.1126/sciadv.aao4364

Vaisvila, R., Ponnaluri, V. K. C., Sun, Z., Langhorst, B. W., Saleh, L., Guan, S., Dai, N., Campbell, M. A., Sexton, B. S., Marks, K., Samaranayake, M., Samuelson, J. C., Church, H. E., Tamanaha, E., Corrêa, I. R., Pradhan, S., Dimalanta, E. T., Evans, T. C., Williams, L., & Davis, T. B. (2021). Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA. *Genome Research*, *31*(7), 1280–1289. https://doi.org/10.1101/gr.266551.120

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Zemach, A., McDaniel, I. E., Silva, P., & Zilberman, D. (2010). Genome-Wide Evolutionary Analysis of Eukaryotic DNA Methylation. *Science*, *328*(5980), 916–919. https://doi.org/10.1126/science.1186366

Zhou, W., Johnson, B. K., Morrison, J., Beddows, I., Eapen, J., Katsman, E., Semwal, A., Habib, W. A., Heo, L., Laird, P. W., Berman, B. P., Triche, T. J., & Shen, H. (2024). BISCUIT: An efficient, standards-compliant tool suite for simultaneous genetic and epigenetic inference in bulk and single-cell studies. *Nucleic Acids Research*, *52*(6), e32–e32. https://doi.org/10.1093/nar/gkae097

Data Accessibility and Benefit-Sharing:

Data Accessibility Statement: All scripts are available at https://github.com/EmilyKerns/VariablePerformanceofBS-SeqandReadMappingSoftware. Sequence data is uploaded to Dryad.

Author Contributions:

E.V.K. and J.N.W. conceived and designed the study and co-wrote the manuscript; E.V.K. wrote the software wrappers, conducted analyses, and produced figures; J.N.W. supervised and funded the work.

Tables and Figures:

Table 1: Search results for different bioinformatic tools associated with bisulfite converted DNA. The search terms used were "software name" AND "DNA methylation," conducted on August 21, 2024.

Tool	**Google Scholar Search Results**	**Web of Science Search Results**
Bismark	6,030	30
BWA meth	313	7
BWA mem	3,570	0
MethylDackel	347	1

Table 2: Population, rearing environment, and sample sizes with associated batch of sequencing for RRBS and WGBS. Abbreviations for populations and rearing environments are consistent with figures in the results.

Population	**Rearing Environment**	**Sample Size *(Sequencing Batch)***
Watson (WT, 60.54, -150.47)	Field (FD)	n=4 *(1)*
	Common Garden (CG)	n=4 *(1)*
		WGBS (n=2) *(1)*
Wik (WK, 60.72, -151.25)	Field	n=4 *(1)*
	Common Garden	n=4 *(1)*
		WGBS (n=2) *(1)*
Roberts (ROB, 50.22, -125.54)	Field	n=3 *(2)*

	Common Garden	n= 2 *(1)*, n = 2 *(2)*
Gosling (GOS, 50.05, -125.50)	Field	n=3 *(2)*
	Common Garden	n=2 *(1)*, n = 2 *(2)*
Sayward (SAY, 50.39, -125.95)	Common Garden	n=2 *(1)*, n = 2 *(2)*

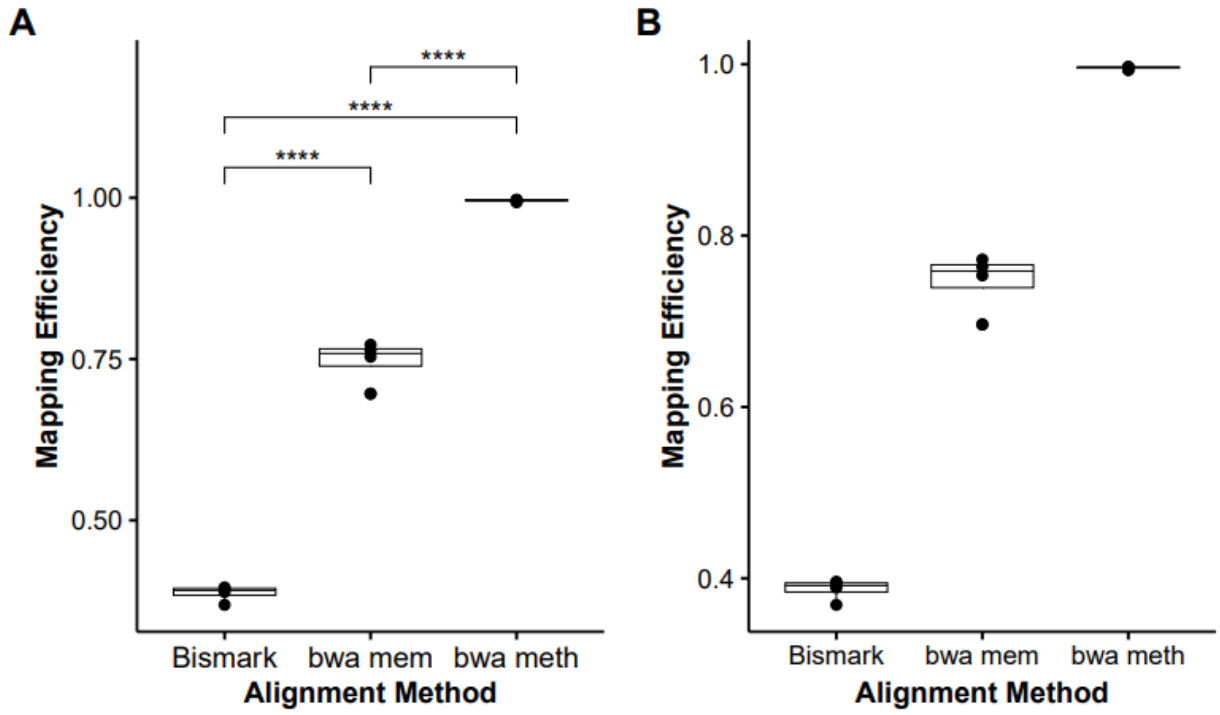

Figure 1: Mapping efficiency of Bismark, BWA mem, and BWA meth on all **(A)** RRBS and **(B)** WGBS samples.

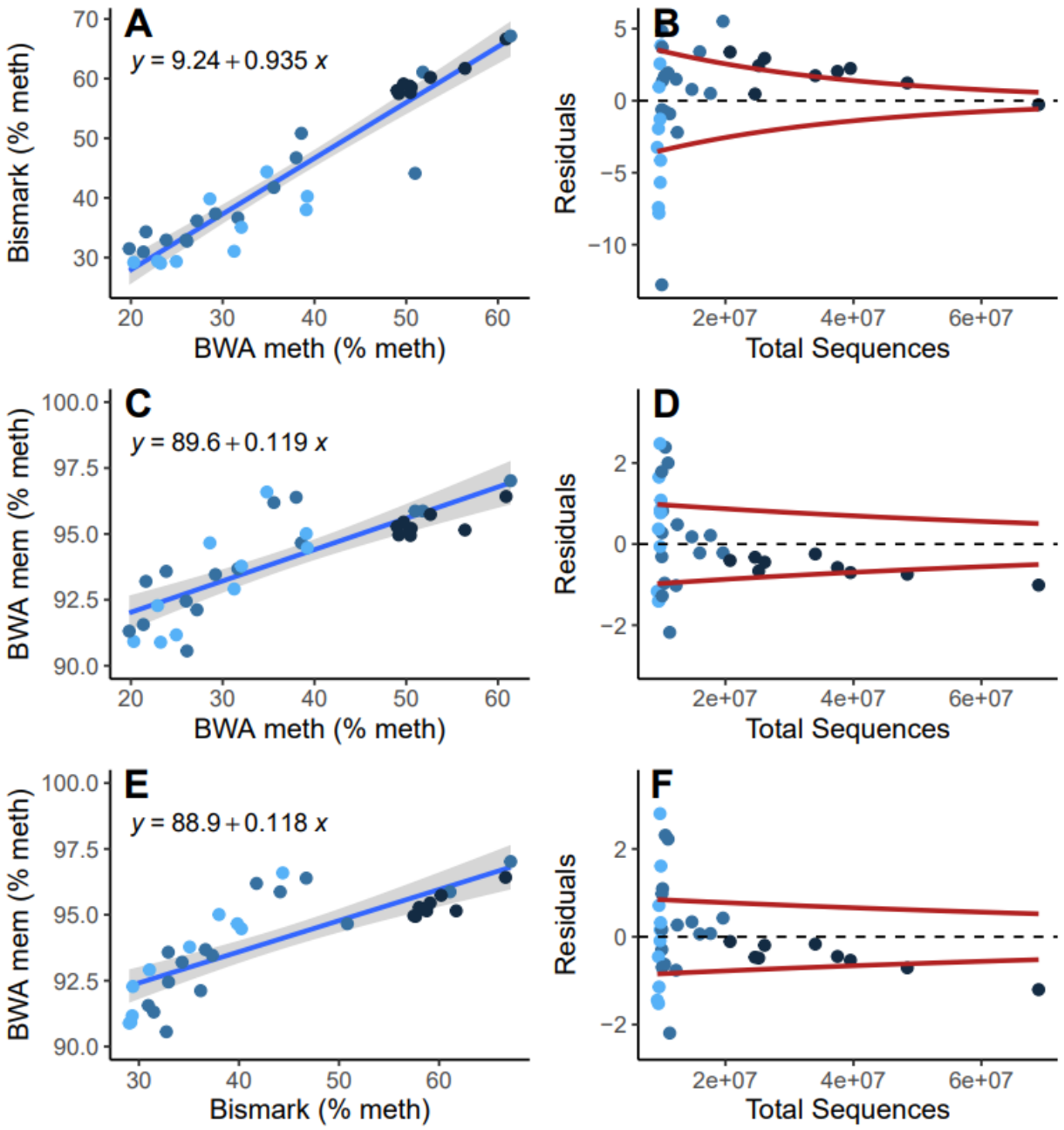

Figure 2: **(A, C, E)** Mean percent methylation per base for each individual sequenced with RRBS using **(A)** BWA meth and Bismark, **(C)** BWA meth and BWA mem, and **(E)** Bismark and BWA mem. A slope that is close to one and y-intercept near zero indicates agreement in mean percent methylation between methods. **(B, D, F)** Total sequences per sample compared to the residuals from the models in the left column. **(B)** For BWA meth versus Bismark, variation in residuals is high when sequencing effort is low, but this variation tends to plateau above 20 million reads per sample. **(D, F)** For both BWA mem comparisons, there isn't a significant relationship between sequencing effort and reduced residuals.

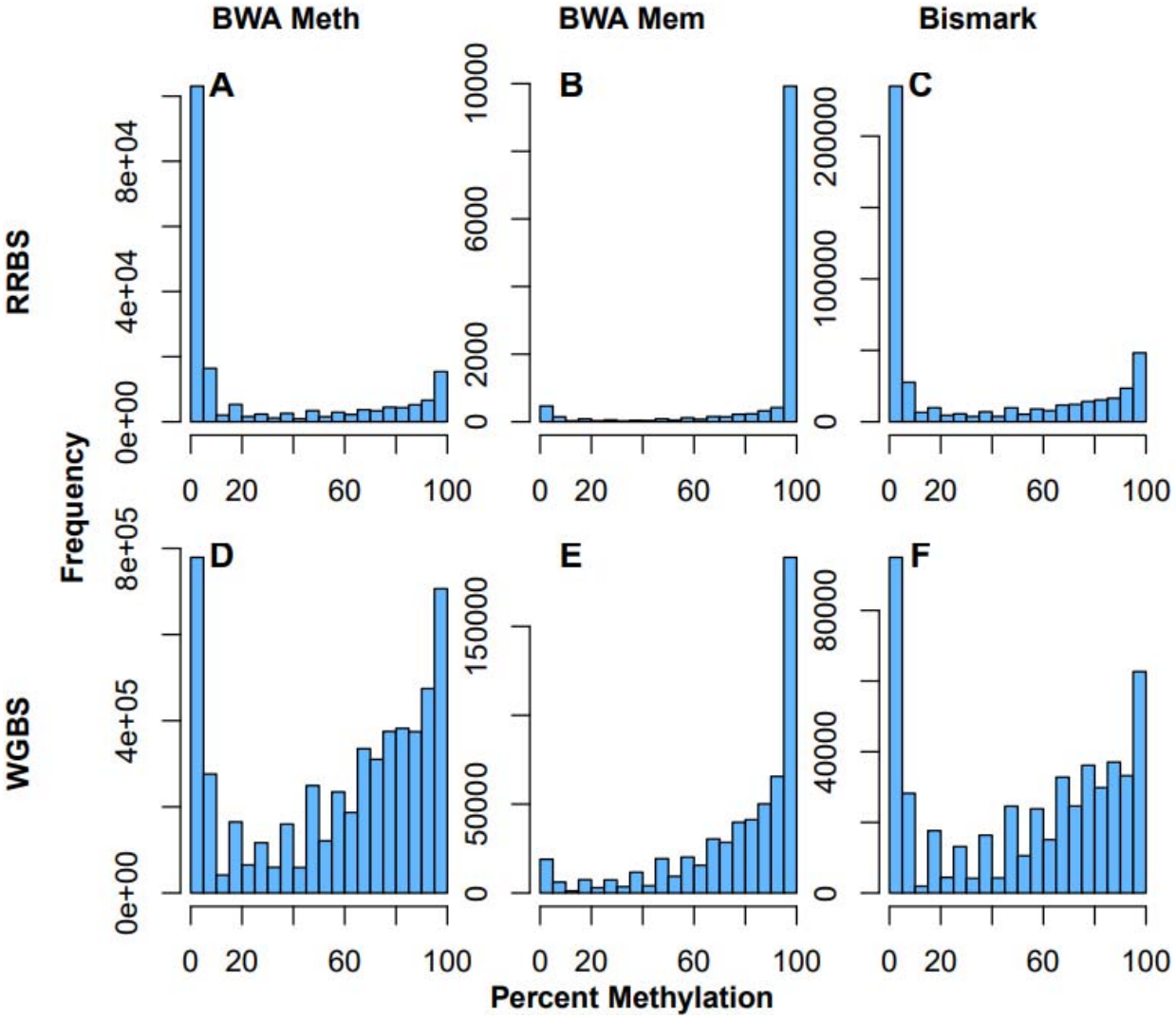

Figure 3: Frequency histograms of percent methylation per base for a technical replicate from Wik lake (sample ID: CG_WK_002). Liver tissue from this sample was sequenced using RRBS (**A-C**) and WGBS (**D-F**) before read mapping with BWA meth (**A, D**), BWA mem (**B, E**), and Bismark (**C, F**). MethylDackel was used for methylation calling and SNP filtering for BWA meth and BWA mem, while there was no SNP filtering for Bismark aligned reads.

Table 3: CpG sites that were in common or unique to WGBS and RRBS for technical replicates with a minimum of 10x depth and 5x depth.

Sample ID	**Minimum Depth**	**Unique to WGBS** ***(Number of CpG sites)***	**Unique to RRBS** ***(Number of CpG sites)***	**Overlapping Sites** ***(Number of CpG sites)***
CG_WK_002	10x	96.6% *(5,417,478)*	2.4% *(134,797)*	1.0% *(55,001)*

	5x	96.0% *(18,501,995)*	1.1% *(219,188)*	2.8% *(546,037)*
CG_WK_003	10x	97.1% *(5,160,689)*	2.2% *(116,016)*	0.7% *(36,879)*
	5x	96.9% *(18,194,034)*	1.0% *(188,716)*	2.1% *(402,195)*
CG_WT_002	10x	98.8% *(4,966,649)*	0.9% *(46,124)*	0.3% *(14,544)*
	5x	97.1% *(17,915,424)*	1.1% *(194,016)*	1.8% *(337,685)*
CG_WT_004	10x	97.1% *(4,410,125)*	2.3% *(102,397)*	0.7% *(31,069)*
	5x	96.5% *(17,312,003)*	1.2% *(209,426)*	2.3% *(418,643)*

Table 4: CpG sites that were in common between two individuals within the using RRBS and WGBS with a minimum of 10x and 5x depth.

Sample Comparison	**Sequencing Method**	**Minimum Depth**	**Unique to First Sample *(Number of CpG sites)***	**Unique to Second Sample *(Number of CpG sites)***	**Overlapping Sites *(Number of CpG sites)***
CG_WT_002 & CG_WT_004	RRBS	10x	22.2% *(38,133)*	64.6% *(110,931)*	13.1% *(22,535)*
		5x	30.4% *(274,939)*	41.1% *(371,307)*	28.4% *(256,762)*
	WGBS	10x	44.0% *(3,493,419)*	37.2% *(2,953,420)*	18.8% *(1,487,774)*
		5x	22.3% *(5,099,302)*	20.0% *(4,576,839)*	57.6% *(13,153,807)*
CG_WK_002 & CG_WK_003	RRBS	10x	43.2% *(116,469)*	29.5% *(79,556)*	27.2% *(73,329)*
		5x	39.0% *(377,396)*	21.0% *(203,082)*	40.1% *(387,829)*
	WGBS	10x	41.7% *(3,715,901)*	38.6% *(3,440,990)*	19.7% *(1,756,578)*
		5x	21.5% *(5,085,066)*	19.6% *(4,633,263)*	59.0% *(13,962,966)*

Table 5: A summary of the limitations and strengths of alignment wrappers available for bisulfite converted DNA. Qualitative descriptions of performance (Low, Moderate, High) are relative to the performance of other mapping tools for bisulfite converted libraries and collated from the results of multiple studies.

Tool	**Algorithm**	**Mapping Efficiency**	**Library Type**	**SNP calling with bisulfite sequencing data**	**Sources**

Bismark (Krueger & Andrews, 2011)	Bowtie2, HiSAT2	Low	Directional & Nondirectional	No	Present study, (Farrell et al., 2021; Guo et al., 2013; Nunn et al., 2021; Zhou et al., 2024)
BWA meth (Pedersen et al., 2014) + **MethylDackel** (https://github.com/dpryan79/MethylDackel)	BWA mem	High	Directional	Yes (can also use BISCUIT in place of MethylDackel)	Present study, (Farrell et al., 2021; Nunn et al., 2021; Zhou et al., 2024)
BISCUIT (Zhou et al., 2024)	BWA mem	High	Directional & Nondirectional	Yes	(Farrell et al., 2021; Zhou et al., 2024)
BiSulfite Bolt (Farrell et al., 2021)	BWA mem	High	Directional & Nondirectional	Yes	(Farrell et al., 2021; Zhou et al., 2024)