



## RESEARCH ARTICLE

# Evaluation of cytosine conversion methods for whole-genome DNA methylation profiling

[version 1; peer review: 2 not approved]

Shyaron Poudel <sup>1-3</sup>, Brett Hale<sup>1-3</sup>, Asela J. Wijeratne <sup>2,3</sup><sup>1</sup>Molecular Biosciences Graduate Program, Arkansas State University, State University, Jonesboro, AR, 72467, USA<sup>2</sup>Arkansas Biosciences Institute, Arkansas State University, State University, Jonesboro, AR, 72467, USA<sup>3</sup>College of Science and Mathematics, Arkansas State University, State University, Jonesboro, AR, 72467, USA

**V1** First published: 07 Dec 2022, 11:1450  
<https://doi.org/10.12688/f1000research.128391.1>  
Latest published: 07 Dec 2022, 11:1450  
<https://doi.org/10.12688/f1000research.128391.1>

## Abstract

**Background:** DNA methylation, the most common epigenetic modification, is defined as the removal or addition of methyl groups to cytosine bases. Studying DNA methylation provides insight into the regulation of gene expression, transposon mobility, genomic stability, and genomic imprinting. Whole-genome DNA methylation profiling (WGDM) is a powerful tool to find DNA methylation. This technique combines standard whole-genome sequencing methodology (e.g., Illumina high-throughput sequencing) with additional steps where unmethylated cytosine is converted to uracil. However, factors such as low cytosine conversion efficiency and inadequate DNA recovery during sample preparation oftentimes render poor-quality data. It is therefore imperative to benchmark sample preparation protocols to increase sequencing data quality and reduce false positives in methylation detection.

**Methods:** A survey analysis was performed to investigate the efficiency of the following commercially available cytosine conversion kits when coupled with the NEBNext® Ultra™ DNA Library Prep Kit for Illumina (NEB): Zymo Research EZ DNA Methylation™ kit (hereafter known as Zymo Conversion kit), QIAGEN EpiTect Bisulfite kit (hereafter known as QIAGEN Conversion kit), and NEBNext® Enzymatic Methyl-seq Conversion Module (hereafter known as NEB EM-seq kit). Input DNA was derived from soybean (*Glycine max* [L.] Merrill) leaf tissue.

**Results:** Of those tested, the QIAGEN Conversion kit provided the best sample recovery and the highest number of sequencing reads, whereas the Zymo Conversion kit had the best cytosine conversion efficiency and the least duplication. The sequence library obtained with the NEB EM-seq kit had the highest mapping efficiency (percentage of reads mapped to the genome). The data quality (defined by Phred score) and methylated cytosine call were similar between kits.

**Conclusions:** This study offers the groundwork for selecting an effective DNA methylation detection kit for crop genome research.

## Open Peer Review

**Approval Status**

	1	2
<b>version 1</b> 07 Dec 2022	 view	 view

1. **Aaron John Stevens** , University of Otago, Wellington, New Zealand
2. **Sadaruddin Chachar** , Sindh Agriculture University, Tando Jam, Pakistan

Any reports and responses or comments on the article can be found at the end of the article.

**Keywords**

DNA methylation, Methylation profiling, Whole-Genome Bisulfite Sequencing (WGBS), Enzymatic methyl-seq (EM-seq), Bisulfite, Cytosine, Uracil



This article is included in the **Genomics and Genetics** gateway.



This article is included in the **Bioinformatics** gateway.



This article is included in the **Bioinformatics Education and Training Collection** collection.

**Corresponding author:** Asela J. Wijeratne ([awijeratne@astate.edu](mailto:awijeratne@astate.edu))

**Author roles:** **Poudel S:** Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Hale B:** Methodology, Writing – Review & Editing; **Wijeratne AJ:** Conceptualization, Formal Analysis, Funding Acquisition, Methodology, Project Administration, Resources, Supervision, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** This work was supported by startup funds from Arkansas BioScience Institute and Arkansas State University to AJW. *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2022 Poudel S *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

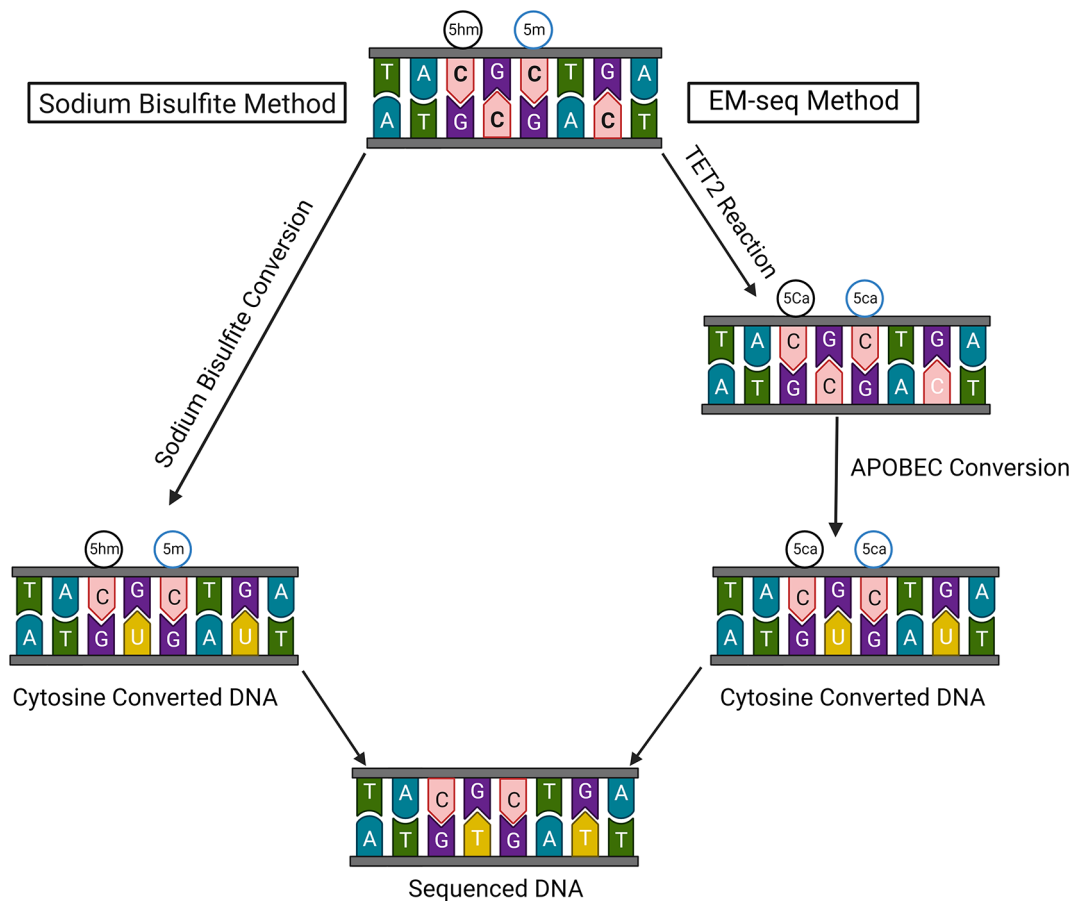
**How to cite this article:** Poudel S, Hale B and Wijeratne AJ. **Evaluation of cytosine conversion methods for whole-genome DNA methylation profiling [version 1; peer review: 2 not approved]** F1000Research 2022, 11:1450  
<https://doi.org/10.12688/f1000research.128391.1>

**First published:** 07 Dec 2022, 11:1450 <https://doi.org/10.12688/f1000research.128391.1>

## Introduction

DNA methylation is a key epigenetic regulator of many biological processes, including transposon activation, heterosis, biotic and abiotic stress response, development, and reproduction (Moore *et al.*, 2013). Whole-genome DNA methylation profiling (WGDM) is among the most efficient techniques available for the detection of DNA methylation, providing >90% single-nucleotide-based genome-wide coverage for cytosines followed by guanine (CpG) residues (Chatterjee *et al.*, 2017). WGDM is similar to the traditional whole-genome sequencing technique with additional steps of cytosine conversion to uracil (Plongthongkum *et al.*, 2014; Li *et al.*, 2018). Of deployed cytosine conversion techniques, sodium bisulfite (SB) conversion is a gold standard process that utilizes single-base-level chemical treatment to convert cytosine to uracil while leaving 5-methylcytosine (5-mC) unaffected prior to PCR enrichment (Figure 1). During PCR amplification, uracil is converted to thiamine, allowing researchers to perform high-throughput sequencing to determine the average methylation level in a sample (Feng *et al.*, 2020). Although considered a gold standard, several limitations can arise through SB treatment, such as low conversion efficiency (failed conversion of non-methylated cytosines to uracil) and DNA degradation. These can impair the efficiency of downstream steps such as PCR amplification, library yield, and false detection of methylated cytosine (Hernández *et al.*, 2013; Iurlaro *et al.*, 2016; Tierling *et al.*, 2018).

An approach known as enzymatic methyl-seq (EM) conversion has recently been developed to circumvent the limitations of SB conversion (Vaisvila *et al.*, 2021). In the first step of this method, the 5-mC is converted to 5-hydroxymethylcytosine (5-hmC), then 5-formylcytosine (5-fC), and eventually to 5-carboxylcytosine (5-caC) using the ten-eleven translocation (TET) enzyme, which protects the methylated cytosines to be converted in downstream steps (Ito *et al.*, 2011). In the second reaction, an enzyme known as apolipoprotein B mRNA editing enzyme catalytic polypeptide-like 3A (APOBEC3A) deaminates non-methylated cytosines into uracil (Figure 1) (Wijesinghe and Bhagwat, 2012; Schutsky *et al.*, 2017). The final sequences containing methylated cytosines remain cytosines, while



**Figure 1. Schematic illustration comparing cytosine conversion steps between WGBS and EM-seq methodologies.** Created with BioRender.com. WGBS, whole-genome bisulfite sequencing; EM-seq, enzymatic methylation sequencing; APOBEC3A, apolipoprotein B mRNA editing enzyme catalytic polypeptide-like 3A.

non-methylated cytosines appear as thymine (like in SB conversion-mediated methylation detection). Therefore, the same analytical methods can be used for both EM- and SB-converted DNA (Feng *et al.*, 2020).

Recent data indicate EM conversion provides superior results compared to SB treatment. In a current research, Feng *et al.*, (2020) investigated DNA methylation in *Arabidopsis thaliana* using the enzymatic methylation sequencing (EM-seq) method and compared their findings to whole-genome bisulfite sequencing (WGBS) data. They reported EM-seq provided better data than that of WGBS. Specifically, libraries prepared with EM-seq kits had higher mapping and lower duplication rates, lower background noise, higher average coverage, and higher total cytosine coverage between replicates (Feng *et al.*, 2020). In another study using human gDNA, similar results were obtained where EM-seq libraries outscored bisulfite-converted libraries in most of the examined parameters, including coverage, duplication, sensitivity, and nucleotide composition (Vaisvila *et al.*, 2021). Additionally, even with extremely low levels of DNA (100 pg), EM-seq proved to be a more reliable method for detecting methylation state than WGBS, which requires 10–200 ng DNA input (Vaisvila *et al.*, 2021).

However, optimized sample preparation protocols are only available for a handful of species and are missing for many important crop plants. In addition, most existing benchmark methods have been performed for mammalian genomes, which are different from plant genomes (*e.g.*, plant genomes are enriched with repetitive elements, as well as CHG and CHH methylation). Therefore, in this study, we have compared two commercial SB conversion kits: Zymo Conversion kit and QIAGEN Conversion kit with an EM conversion kit, NEB EM-seq kit, using soybean (*Glycine max* [L.] Merrill) DNA samples. Further, we have benchmarked the use of these three kits with the NEBNext<sup>®</sup> Ultra™ DNA Library Prep Kit for Illumina (NEB) for sample preparation for high-throughput sequencing.

## Methods

### Plant growth and DNA isolation

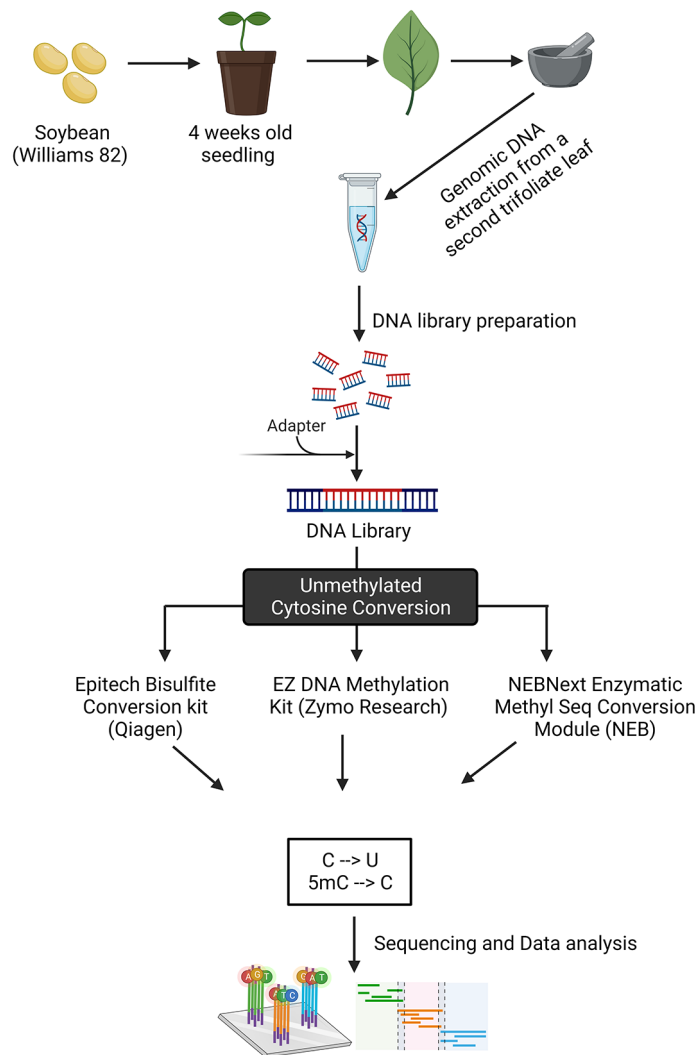
Seeds of soybean genotype Williams 82 were sown in a Miracle-Gro Moisture Control potting medium and grown at 25°C, 16 h d<sup>-1</sup> light at 230 to 365 μM m<sup>-2</sup> s<sup>-1</sup>, and 90% relative humidity in a growth chamber. At the V2 developmental stage (Fehr and Caviness, 1977), a trifoliate leaf was collected for DNA isolation (Figure 2).

A detailed protocol can be found on [protocols.io](https://www.protocols.io). Genomic DNA (gDNA) was isolated using the Zymo Research *Quick-DNA Plant/Seed Miniprep* kit (Cat #D6020; Irvine, CA, USA). DNA concentration was measured using a Qubit fluorometer (Thermo Fisher Scientific; Waltham, MA, USA) coupled with a Thermo Fisher Scientific dsDNA High Sensitivity Assay kit (Cat #Q32851). DNA purity was assessed from 260/230 and 260/280 nm absorbance ratios using a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific). Furthermore, the quality and size of obtained DNA was determined using gel electrophoresis (Fisher Scientific gel casting and electrophoresis apparatus (Cat #FB-SB-1316); 1% agarose gel with 1X Tris acetate-EDTA buffer (ThermoFisher; Cat # B49). Nucleic Acid was stained with Biotium GelRed Nucleic Acid Gel Stain (Fisher Scientific; Cat # NC9594719) and gel imaging was performed using LiCOR ODYSSEY imager (Serial # OFC1321) under a 520 nm channel. As a non-methylated control, the soybean gDNA was spiked with *Escherichia coli* gDNA (Zymo Research, Cat #D5016) representing 1% of the total gDNA input. The spiked sample was then separated into three aliquots (each 65 μL), which were sheared using Qsonica sonicators (Newtown, CT, USA). Sonication settings entailed 15 sec on/90 sec off for 8 cycles at a 20 kHz frequency, resulting in DNA fragments of approximately 350 bp. Sheared aliquots were used for library preparation, as described below.

### Whole-genome Illumina sequencing library preparation

Sequencing libraries were prepared using the NEBNext<sup>®</sup> Ultra™ II DNA Library Prep Kit for Illumina<sup>®</sup> (New England Biolabs; Cat #E7645S; Ipswich, MA, USA). A total of 200 ng (for NEB EM-seq kit) or 1 μg (Zymo and QIAGEN Conversion kits) of starting DNA was used based on the manufacturer's instructions (Figure 2). The sheared DNA fragments were repaired, and Illumina Methylated Adaptors were ligated using NEBNext<sup>®</sup> Multiplex Oligos for Illumina (NEB Cat #E7535S/L). Adapter-ligated fragments were then purified using Solid Phase Reversible Immobilization (SPRI) magnetic beads (Beckman Coulter Inc.; Cat #B23317; Brea, CA, USA). For the cytosine conversion, the adapter-ligated, purified DNA samples were processed with either a Zymo Conversion kit (EZ DNA Methylation™ Kit, Zymo Research; Cat #D5001), QIAGEN Conversion kit (EpiTect Bisulfite Kit, QIAGEN; Cat #59104), or NEB EM-seq kit (NEBNext<sup>®</sup> Enzymatic Methyl-seq Conversion Module, New England Biolab; Cat #E7125S/L) (Figure 2).

For SB conversion using the Zymo Conversion kit, DNA was incubated with CT Conversion Reagent (Zymo Research; Cat #D5001-1) (Table 3 and Figure 1) in a thermocycler for approximately 16 hrs. The converted DNA samples were then desulphonated using M-Desulphonation Buffer (Zymo Research; Cat #D5001-5). For SB conversion using the QIAGEN Conversion kit, the sample was incubated in a thermocycler for approximately 5 hrs with the Bisulfite mix (Table 2 and Figure 1) and the DNA protection buffer provided in the kit. Converted DNA samples were then treated with Buffer BD



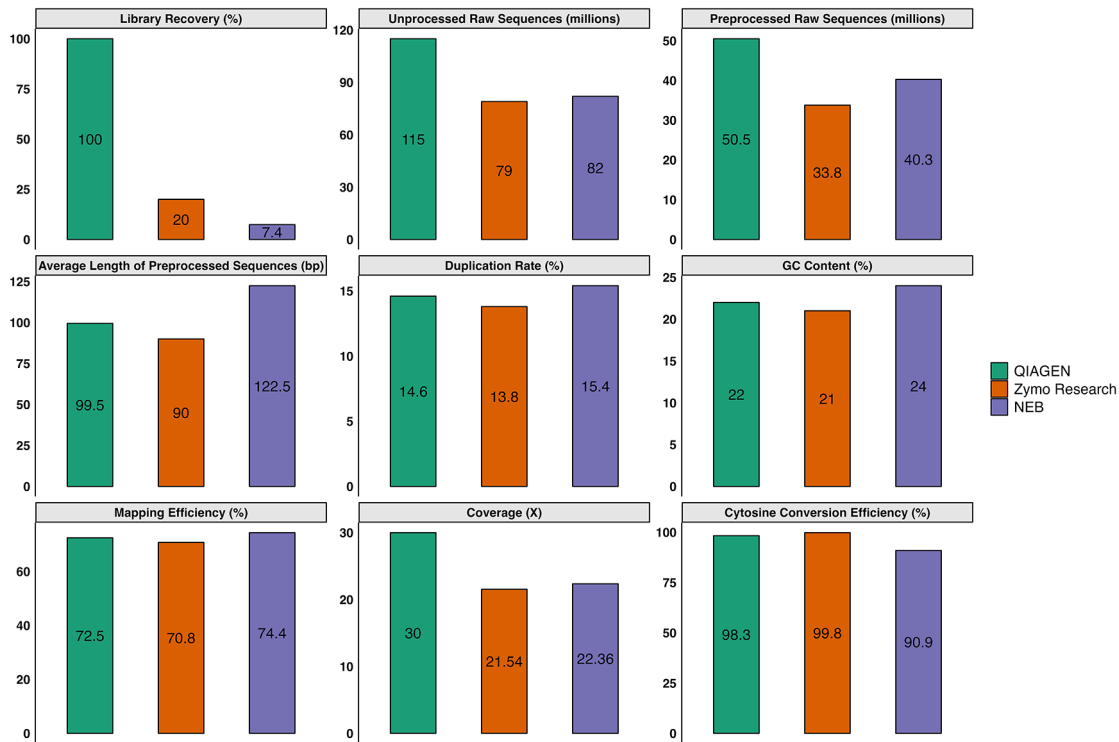
**Figure 2. Schematic illustration of experimental design and project approach for WGBS and EM-seq library preparation.** Created with [BioRender.com](https://BioRender.com). WGBS, whole-genome bisulfite sequencing; EM-seq, enzymatic methylation sequencing; NEB, New England BioLabs.

from the kit for desulphonation. For EM conversion, the sample was first treated with TET2 (NEB Cat #E7130AVIAL; part of the NEBNext<sup>®</sup> Enzymatic Methyl-seq Conversion Module) and APOBEC (NEB Cat #E7133AVIAL) (Table 1 and Figure 1) enzymes in a series of steps, which ultimately converted cytosines to uracil, allowing for the detection of 5 mC and 5 hMC.

Converted samples were purified using Beckman Coulter SPRIselect magnetic beads (Cat #B23317). The barcoding of converted DNA fragments was carried out with NEBNext<sup>®</sup> Multiplex Oligos for Illumina (Methylated Adaptor, Index Primers Set 1) (NEB Cat #E7535S/L) using PCR (16 cycles) with EpiMark<sup>®</sup> Hot Start *Taq* DNA Polymerase (NEB Cat #M0490S) following manufacturer instructions (more details can be found on [protocols.io](https://protocols.io)).

### Quality control and sequencing of WGBS and EM-seq libraries

After amplification, prepared libraries were purified using SPRIselect magnetic beads (Cat #B23317). The concentration of prepared libraries was measured using the Thermo Fisher Scientific dsDNA High Sensitivity Assay kit (Cat #Q32851) with an Invitrogen Qubit Fluorometer as well as a NanoDrop ND-1000 spectrophotometer (Cat# E1123552; Thermo Fisher Scientific). Prepared libraries were then sent to Novogene Corporation Inc. (Sacramento, CA, USA) for downstream quality assessment and whole-genome sequencing. Fragment size and concentration were validated with an Agilent Bioanalyzer (Agilent Technologies; Santa Clara, CA, USA). Libraries were sequenced on an Illumina Hi-Seq sequencer to obtain 150 bp paired-end reads.



**Figure 3. Comparison of sequencing data quality between the cytosine conversion kits.** Quality was determined by library recovery (%), number of unprocessed and preprocessed raw sequences, average length of preprocessed sequences (bp), duplication rate (%), GC content (%), mapping efficiency (%), total coverage (X), and cytosine conversion efficiency (%). NEB, New England BioLabs.

### Data analysis

The quality of raw reads obtained from the three samples was assessed using FastQC (RRID:SCR\_014583) (v.0.11.8) (Andrews, 2010) and MultiQC (RRID:SCR\_014982) (v1.9) (Ewels *et al.*, 2016). Adapter sequences and poor-quality bases (Phred <20) were removed using Trim Galore (RRID:SCR\_011847) (v0.4.2) (Bolger *et al.*, 2014). The pre-processed reads were then aligned to the soybean reference genome (Gmax\_275\_v2.0) obtained from Phytozome (Goodstein *et al.*, 2012) using Bismark (RRID:SCR\_005604) (v.0.22.3) aligner with default parameters (Krueger and Andrews, 2011). After removing duplicate reads, Bismark (v.0.22.3) (Krueger and Andrews, 2011) was used to detect cytosine methylation at a single-base resolution. Cytosine methylation levels at each region of the sequence were calculated as the number of methylated C vs. total C and T present. To estimate and evaluate false-positive methylation levels and efficiency of cytosine deamination, reads were aligned to the non-methylated controls, soybean chloroplast genome (NC\_007942.1), and *E. coli* non-methylated genomic DNA sequences. Data were visualized with RStudio (RRID:SCR\_000432) (v1.1.463) (R Core Team, 2020) implementing the ggplot2 package (RRID:SCR\_014601) (v.3.3.5) (Wickham, 2016). The code is available from GitHub and is archived with Zenodo (Wijeratne, 2022).

### Results and discussion

Despite the utility and growing popularity of WGBS for epigenome analysis in crop plants, relatively little has been done to identify various factors affecting sample preparation and data quality. Cytosine conversion is a crucial step of this process, as partial conversion can lead to false methylation calls. In this survey experiment, we used soybean DNA to prepare samples using three different cytosine conversion protocols. We have evaluated their ease of use, DNA loss during the conversion step, and subsequent data quality. These evaluations shed light on the performance of these three kits and their use for preparing samples to analyze crop plant epigenomes.

#### The QIAGEN kit outperformed other kits in time requirement and DNA recovery

Essential factors to consider when selecting a kit are the final yield of libraries and the ease of the protocol. Based on the time required for cytosine conversion steps, the QIAGEN Conversion kit was the fastest method to prepare the whole library ( $\approx 24$  hrs), followed by the NEB EM-seq kit ( $\approx 26$  hrs) and Zymo Conversion kit ( $\approx 35$  hrs). When evaluating the cytosine conversion step, the Zymo Conversion kit needed the longest time of  $\approx 16$  hrs. However, the QIAGEN

**Table 1. Thermocycler incubation condition for the NEB kit.** NEB, New England BioLabs.

NEBNext Enzymatic Methyl Seq Conversion Module (NEB)				
Step	Time	Temperature	Cycles	Total time required for cytosine conversion
Oxidation of 5-methylcytosines and 5-hydroxymethylcytosines	1 hr 30 min	37°C	1	≈ 5 hr.
Denaturation of DNA	10 min	50°C		
Deamination of cytosines	3 hr	37°C		
Hold	∞	4°C		

**Table 2. Thermocycler incubation condition for the Qiagen kit.**

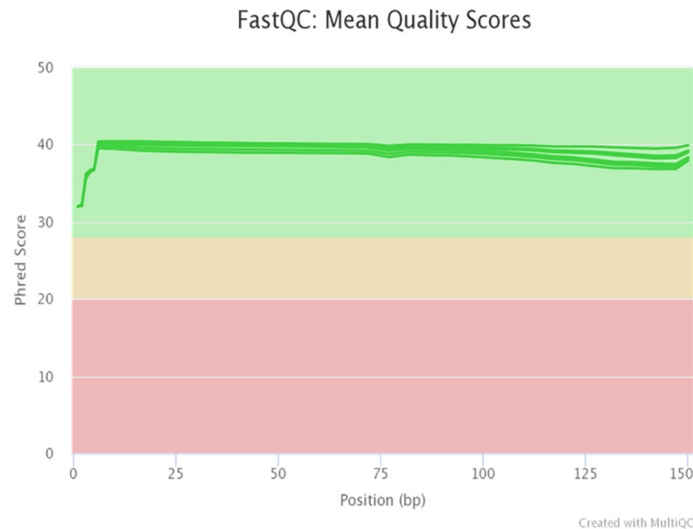
EpiTech Bisulfite Conversion kit (Qiagen)				
Step	Time	Temperature	Cycles	Total time required for cytosine conversion
Denaturation	5 min	95°C	1	≈ 5 hr
Incubation	25 min	60°C		
Denaturation	5 min	95°C		
Incubation	1 hr 25 min	60°C		
Denaturation	5 min	95°C		
Incubation	2 h 55 min	60°C		
Hold	∞	20°C		

**Table 3. Thermocycler incubation condition for the Zymo Research kit.**

EZ DNA Methylation Kit (Zymo Research)				
Step	Time	Temperature	Cycles	Total time required for cytosine conversion
Denaturation	30 sec	95°C	55	≈ 16 hr
Bisulfite conversion	15 min	50°C		
Hold	∞	4°C		

Conversion kit and the NEB EM-seq kit required a total of ≈ 5 hrs, which included the post-conversion cleanup step (Tables 1-3).

Following cytosine conversion, the highest recovery of DNA was observed with the QIAGEN conversion kit (100%), followed by the Zymo Conversion (20%) and NEB EM-seq (7.40%) kits (Figure 3). Previously, researchers noted factors that resulted in the reduction of libraries while using SB-based cytosine conversion. In a study conducted by Tanaka and Okamoto (2007), a real-time PCR experiment indicated that DNA degradation during SB conversion occurs primarily due to the hydrolytic reaction or the prolonged exposure to the SB. Additionally, prolonged exposure to SB results in the formation of abasic sites and subsequent DNA strand damage (Tanaka and Okamoto, 2007). An abasic site is a region of DNA that lacks both purine and pyrimidine bases due to DNA damage formed by spontaneous hydrolysis of the N-glycosidic bond (Boiteux and Guillet, 2004). According to the manufacturer's manuals for SB-based conversion kits, DNA is denatured and fragmented before or during cytosine conversion in the thermocycler to chemically convert non-methylated cytosines into uracil. Thus, the prolonged exposure to SB during denaturation, as well as DNA fragmentation, likely contributed to the loss of small DNA fragments during the downstream purification steps, reducing library yield. Conversely, high recovery of DNA when using the QIAGEN conversion kit may be attributed to the presence of DNA protection buffer in the kit (Izzi *et al.*, 2014; Leontiou *et al.*, 2015; Tierling *et al.*, 2018). The drastic loss of DNA with the NEB EM-seq kit may be due to comparatively more purification steps, increasing the chances of DNA loss due to pipetting.



**Figure 4. Mean quality scores of WGBS and EM-seq libraries.** The result was obtained using MultiQC. WGBS, whole-genome bisulfite sequencing; EM-seq, enzymatic methylation sequencing.

#### All cytosine conversion kits yielded quality sequencing data

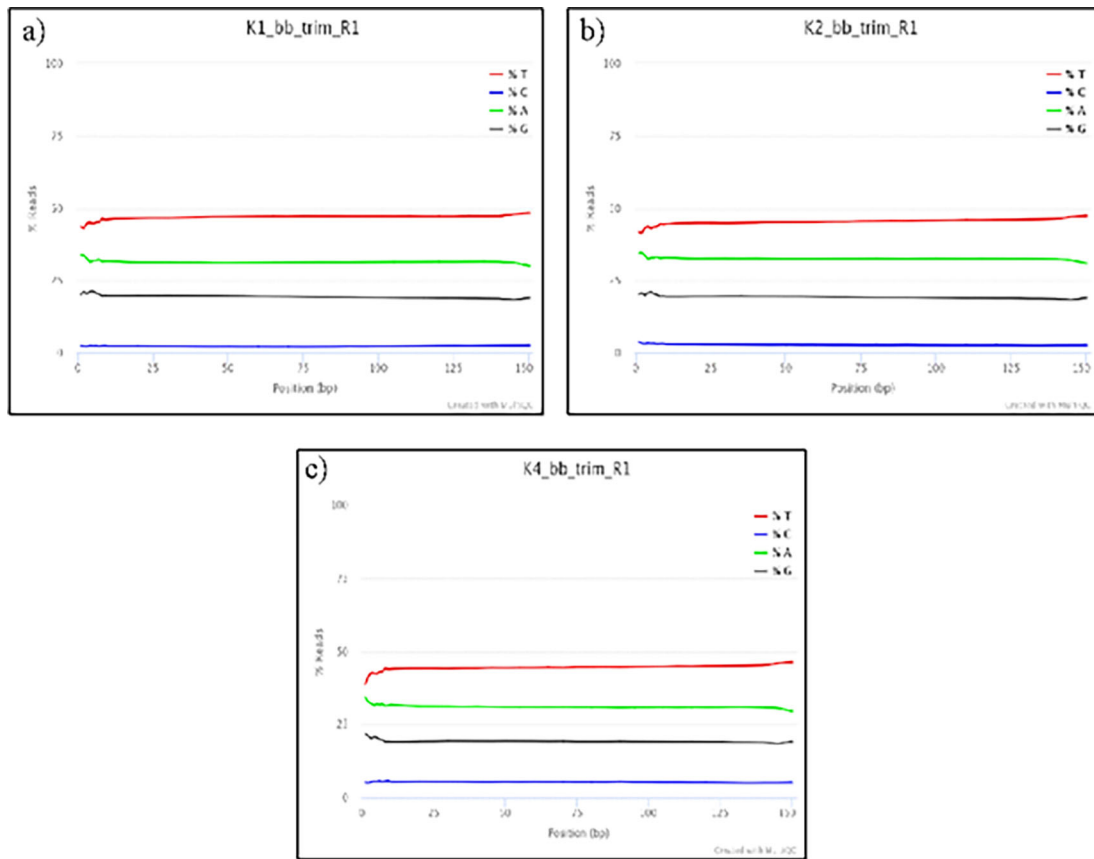
Prepared samples were sequenced to see if the cytosine conversion method had any effect on the sequencing. The initial quality of sequencing data was analyzed based on the following criteria: the total number of sequencing reads obtained, total coverage (X), duplication rate (%), Phred quality score, Per-base sequence, and GC content (%) in the sample. These parameters indicated no major differences among the three samples as discussed below. The number of raw sequencing reads per sample ranged from 33–50 million, with coverage ranging from 20X to 30X (Figure 3). The Phred quality score of each data set was found to be excellent, ranging between 31 to 40 (Figure 4). Based on the per-base sequence plot, thymine was the most abundant base in each library with 50% of all four bases, whereas cytosine was the least abundant (Figure 5). GC content was similar between samples, with the Zymo Research library having the highest level, followed by the QIAGEN and NEB kits (Figure 3). As expected, the per-sequence GC content for each sample was biased due to cytosine conversion into thiamine during PCR amplification. Thus, our data suggest that the cytosine conversion method has little effect on the initial sequence quality.

Nevertheless, initial data quality may not reflect data utility. Therefore, we mapped reads to the soybean genome to see the mapping efficiency (defined as how many reads can be aligned to the genome proportional to the total reads). Reads generated from all three methods had similar mapping efficiencies. The NEB kit library had the highest mapping efficiency (74.4%), followed by the QIAGEN kit (72.5%) and the Zymo Research kit (70.8%) (Figure 3). The observed mapping efficiency (>70%) is comparable with previously reported efficiencies for cytosine-converted samples ( $\approx$  60–75%) (Hari and Parthasarathy, 2019). These results implied that sequences obtained from all three methods could be aligned correctly to the soybean genome.

We also calculated the duplication rate for each sample, as a higher rate of duplication distorts the true sequence proportion in a given sample. Duplication rates between 13 and 15% were detected for each library (Figure 3), which is lower than previously reported duplication rates during soybean methylome profiling (Rambani *et al.*, 2015). Duplication can occur for several reasons. For instance, biased PCR enrichment and overamplification of DNA fragments can result in an overrepresentation of library fragments. Duplication can also occur when an identical template binds to numerous clusters on a flow cell. Duplicates of PCR products distort the actual proportion of sequences in the sample (please see more details on the [Babraham Institute website](#)).

#### Cytosine conversion efficiency

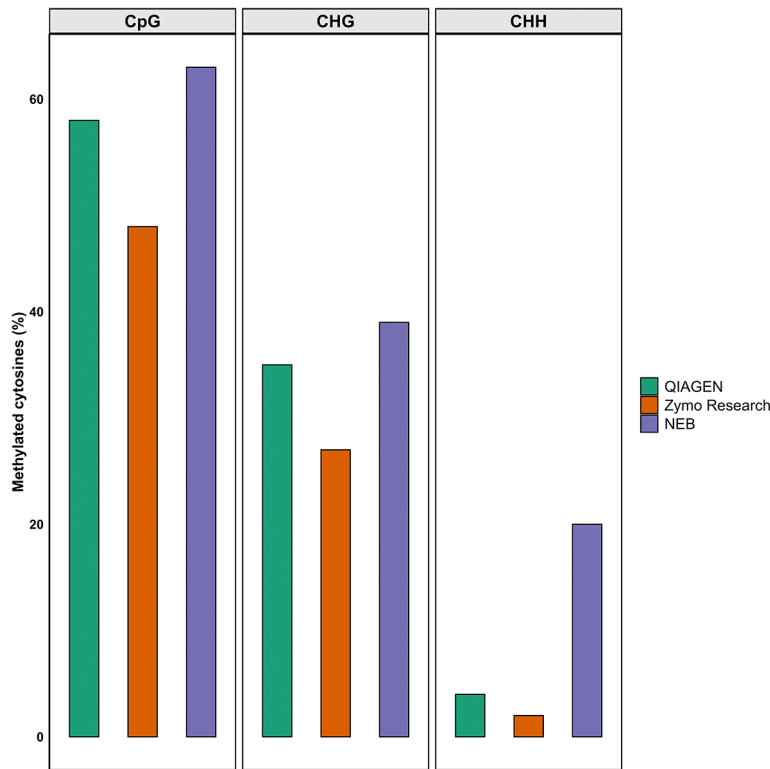
It is crucial to analyze the cytosine conversion efficiency during DNA methylation analysis. If a library has low conversion efficiency, it can result in false methylation detection (Singer, 2019). Adding methylated, non-methylated, or both types of genomic controls in a sample can help identify these limitations presented by bisulfite-converted DNA. Here, the cytosine conversion efficiency was calculated using reads aligned to non-methylated *E. coli* and chloroplast gDNA for each sample. According to these alignments, the bisulfite conversion efficiency of samples ranged between 90 and 99.8% (Figure 3). QIAGEN and Zymo Conversion kits had a similar conversion efficiency of roughly 99%



**Figure 5. Per Base Sequence Content of the (a) Zymo-research kit, (b) Qiagen kit, and (c) NEB kit.** Results were obtained using MultiQC. NEB, New England BioLabs.

(Figure 3). Similar results were reported for both bisulfite kits in a study conducted on human chromosomal DNA (Tierling *et al.*, 2018). For the NEB EM-seq kit, the cytosine conversion efficiency was found to be 90.9%, around 9% lower than the data obtained from a published study (Figure 3) (Feng *et al.*, 2020). The efficiency and accuracy of a conversion kit are dependent primarily on the PCR cycling procedure, conditions, the length of the desulphonation process, and the addition of reagents to prevent DNA degradation (Izzi *et al.*, 2014; Tierling *et al.*, 2018). In this case, the QIAGEN kit included a DNA protection buffer containing a pH indicator dye to ensure the correct pH for cytosine conversion. Alternatively, the Zymo Research kit protocol distinguishes the alkaline or denaturation step from the conversion step (Izzi *et al.*, 2014). However, it is unclear what contributed to poor cytosine conversion efficiency in the NEB EM-seq kit. It is possible that further optimization of the input DNA quantity is necessary for this kit when it is used with the NEB library preparation kit.

Furthermore, the data were aligned to the soybean reference genome to detect the number of methylated cytosines in each library. The highest number of methylated cytosines in all three contexts (*i.e.*, CpG, CHG, and CHH) was found in the library prepared with the NEB EM-seq kit, with around  $1.295 \times 10^9$  total methylated cytosines, followed by  $1.169 \times 10^9$  in the QIAGEN library and 703 million in the Zymo Research library. Moreover, the QIAGEN and Zymo Research libraries displayed a similar distribution of methylated cytosine across contexts. The NEB library showed a distinct distribution compared to the other kits, with an increase in methylated cytosines at CHH context (Figure 6). This can be attributed to the comparatively poor cytosine conversion efficiency, which may have resulted in a high number of erroneously methylated cytosines (Simpson *et al.*, 2017). As expected, the majority of cytosines in each library were methylated in CpG context, followed by CHG and CHH (Figure 6). These data are congruent with a past soybean DNA methylation study, which revealed the greatest levels of methylation in CpG context, followed by CHG and CHH contexts, under normal developmental conditions (Song *et al.*, 2013).



**Figure 6.** Percentage of methylated cytosines detected in CpG, CHG, and CHH methylation contexts in sequencing data obtained from libraries prepared using the Zymo-research kit, Qiagen kit, and NEB kit. NEB, New England BioLabs.

## Conclusions

The current work is the first to compare WGBS and EM-seq to analyze DNA methylation in the soybean genome. Here, a survey study was conducted deploying three commercially available DNA methylation detection kits. The results obtained suggested that the QIAGEN kit provided the best DNA yield and number of sequencing reads. The Zymo Research kit demonstrated the best cytosine conversion efficiency and a low duplication rate. Nevertheless, we recommend the NEBNext<sup>®</sup> Enzymatic Methyl-seq Conversion Module, which was superior to the other kits based upon required DNA input, quality of generated libraries, and hands-on time. Follow-up studies with biological/technical replication are needed to validate the reproducibility and efficiency of the defined conversion kits.

## Data availability

### Underlying data

Raw sequencing data are available as follows:

BioProject: *Glycine max* cultivar: Williams 82 (soybean). Accession number PRJNA902392; <https://identifiers.org/bioproject:PRJNA902392> (Arkansas State University, 2022a).

Sequence Read Archive: Evaluation of cytosine conversion methods for whole-genome DNA methylation profiling (Run: SRR22331514). Accession number SRX18304416; <https://identifiers.org/insdc.sra:SRX18304416> (Arkansas State University, 2022b).

Sequence Read Archive: Evaluation of cytosine conversion methods for whole-genome DNA methylation profiling (Run: SRR22331515). Accession number SRX18304415; <https://identifiers.org/insdc.sra:SRX18304415> (Arkansas State University, 2022c).

Sequence Read Archive: Evaluation of cytosine conversion methods for whole-genome DNA methylation profiling (Run: SRR22331516). Accession number SRX18304414; <https://identifiers.org/insdc.sra:SRX18304414> (Arkansas State University, 2022d).

## Extended data

Analysis code available from: [https://github.com/ajwije/DNA\\_methylation\\_analysis.git](https://github.com/ajwije/DNA_methylation_analysis.git)

Archived analysis code at time of publication: <https://doi.org/10.5281/zenodo.7328525> (Wijeratne, 2022).

License: MIT

## References

- Andrews S: **FastQC: a quality control tool for high throughput sequence data.** 2010.
- Arkansas State University: *Glycine max* cultivar: Williams 82 (soybean). [Dataset]. *BioProject*. 2022a. [Reference Source](#)
- Arkansas State University: Evaluation of cytosine conversion methods for whole-genome DNA methylation profiling (Run: SRR22331514). [Dataset]. *Sequence Read Archive*. 2022b. [Reference Source](#)
- Arkansas State University: Evaluation of cytosine conversion methods for whole-genome DNA methylation profiling (Run: SRR22331515). [Dataset]. *Sequence Read Archive*. 2022c. [Reference Source](#)
- Arkansas State University: Evaluation of cytosine conversion methods for whole-genome DNA methylation profiling (Run: SRR22331516). [Dataset]. *Sequence Read Archive*. 2022d. [Reference Source](#)
- Boiteux S, Guillet M: **Abasic sites in DNA: repair and biological consequences in *Saccharomyces cerevisiae*.** *DNA Repair*. 2004; **3**: 1–12. [PubMed Abstract](#) | [Publisher Full Text](#)
- Bolger AM, Lohse M, Usadel B: **Trimmomatic: A flexible trimmer for Illumina sequence data.** *Bioinformatics*. 2014; **30**(15): 2114–2120. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chatterjee A, Rodger EJ, Morison IM, *et al.*: **Tools and Strategies for Analysis of Genome-Wide and Gene-Specific DNA Methylation Patterns.** Seymour GJ, Cullinan MP, Heng NCK, editors. *Oral Biology: Molecular Techniques and Applications*. Springer; 2017; (pp. 249–277). [Publisher Full Text](#)
- Ewels P, Magnusson M, Lundin S, *et al.*: **MultiQC: Summarize analysis results for multiple tools and samples in a single report.** *Bioinformatics*. 2016; **32**(19): 3047–3048. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Fehr WR, Caviness CE: **Stages of soybean development. Special Report 80, Iowa Agricultural Experiment Station, Iowa Cooperative External Series, Iowa State University, Ames.** 1977.
- Feng S, Zhong Z, Wang M, *et al.*: **Efficient and accurate determination of genome-wide DNA methylation patterns in Arabidopsis thaliana with enzymatic methyl sequencing.** *Epigenetics & Chromatin*. 2020; **13**(1): 42. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Goodstein DM, Shu S, Howson R, *et al.*: **Phytozome: A comparative platform for green plant genomics.** *Nucleic Acids Research*. 2012; **40**(D1): D1178–D1186. [Publisher Full Text](#)
- Hari R, Parthasarathy S: **Next Generation Sequencing Data Analysis. Encyclopedia of Bioinformatics and Computational Biology.** *ABC of Bioinformatics*. 2019; **1–3**: 157–163.
- Hernández HG, Tse MY, Pang SC, *et al.*: **Optimizing methodologies for PCR-based DNA methylation analysis.** *BioTechniques*. 2013; **55**(4): 181–197. [PubMed Abstract](#) | [Publisher Full Text](#)
- Ito S, Shen L, Dai Q, *et al.*: **Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine.** *Science*. 2011; **333**: 1300–1303. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Iurlaro M, McInroy GR, Burgess HE, *et al.*: **In vivo genome-wide profiling reveals a tissue-specific role for 5-formylcytosine.** *Genome Biology*. 2016; **17**(1): 141. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Izzi B, Binder AM, Michels KB: **Pyrosequencing Evaluation of Widely Available Bisulfite Conversion Methods: Considerations for Application.** *Medical Epigenetics*. 2014; **2**(1): 28–36. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Krueger F, Andrews SR: **Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications.** *Bioinformatics*. 2011; **27**(11): 1571–1572. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Leontiou CA, Hadjidaniel MD, Mina P, *et al.*: **Bisulfite Conversion of DNA: Performance Comparison of Different Kits and Methylation Quantitation of Epigenetic Biomarkers that Have the Potential to Be Used in Non-Invasive Prenatal Testing.** *PLoS One*. 2015; **10**(8): e0135058. [Publisher Full Text](#)
- Li Q, Hermanson PJ, Springer NM: **Detection of DNA Methylation by Whole-Genome Bisulfite Sequencing.** Lagrimini LM, editor. *Maize: Methods and Protocols*. Springer; 2018; (pp. 185–196). [Publisher Full Text](#)
- Moore LD, Le T, Fan G: **DNA Methylation and Its Basic Function.** *Neuropsychopharmacology*. 2013; **38**(1): 23–38. [Publisher Full Text](#)
- Plongthongkum N, Diep DH, Zhang K: **Advances in the profiling of DNA modifications: Cytosine methylation and beyond.** *Nature Reviews Genetics*. 2014; **15**(10): 647–661. [Publisher Full Text](#)
- Rambani A, Rice JH, Liu J, *et al.*: **The Methylome of Soybean Roots during the Compatible Interaction with the Soybean Cyst Nematode.** *Plant Physiology*. 2015; **168**(4): 1364–1377. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- RStudio Team: *RStudio: Integrated Development for R*. PBC, Boston, MA: RStudio; 2020. [Reference Source](#)
- Schutsky EK, Nabel CS, Davis AKF, *et al.*: **APOBEC3A efficiently deaminates methylated, but not TET-oxidized, cytosine bases in DNA.** *Nucleic Acids Research*. 2017; **45**(13): 7655–7665. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Simpson JT, Workman RE, Zuzarte PC, *et al.*: **Detecting DNA cytosine methylation using nanopore sequencing.** *Nature Methods*. 2017; **14**(4): 407–410. [PubMed Abstract](#) | [Publisher Full Text](#)
- Singer BD: **A Practical Guide to the Measurement and Analysis of DNA Methylation.** *American Journal of Respiratory Cell and Molecular Biology*. 2019; **61**(4): 417–428. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Song Q-X, Lu X, Li Q-T, *et al.*: **Genome-Wide Analysis of DNA Methylation in Soybean.** *Molecular Plant*. 2013; **6**(6): 1961–1974. [PubMed Abstract](#) | [Publisher Full Text](#)
- Tanaka K, Okamoto A: **Degradation of DNA by bisulfite treatment.** *Bioorganic & Medicinal Chemistry Letters*. 2007; **17**(7): 1912–1915. [Publisher Full Text](#)
- Tierling S, Schmitt B, Walter J: **Comprehensive Evaluation of Commercial Bisulfite-Based DNA Methylation Kits and Development of an Alternative Protocol With Improved Conversion Performance.** *Genetics & Epigenetics*. 2018; **10**: 1179237X18766097. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Vaisvila R, Ponnaluri VKC, Sun Z, *et al.*: **Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA.** *Genome Research*. 2021; **31**(7): 1280–1289. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wickham H: *ggplot2: Elegant Graphics for Data Analysis*. Springer; 2016.
- Wijeratne A: **ajwije/DNA\_methylation\_analysis: v1.0.0 (v1.0.0).** *Zenodo*. [Code]. 2022. [Publisher Full Text](#)
- Wijesinghe P, Bhagwat AS: **Efficient deamination of 5-methylcytosines in DNA by human APOBEC3A, but not by AID or APOBEC3G.** *Nucleic Acids Research*. 2012; **40**(18): 9206–9217. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

# Open Peer Review

Current Peer Review Status:  

---

Version 1

Reviewer Report 04 April 2023

<https://doi.org/10.5256/f1000research.140976.r167362>

© 2023 Chachar S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Sadaruddin Chachar** 

Sindh Agriculture University, Tando Jam, Pakistan

Overall, this research article represents an interesting investigation on “Evaluation of cytosine conversion methods for whole-genome DNA methylation profiling”, to assess the performance of various DNA methylation profiling techniques and evaluate their validity for high-throughput sequencing. Abstract seems logical providing the concise summary of the findings, the introduction provides sufficient background of the study, while the methods are not generally appropriate for the experiments conducted. The analysis and results presented in figures seem logical while interpretation is supported by results. Moreover, the results are clearly described, making the manuscript easily understandable for readers. All the comments and remarks are given below.

In the first line of abstract, the correct definition of DNA methylation is the addition of a methyl group to the cytosine base in the DNA molecule, not the removal or addition of methyl groups to cytosine bases.

Discuss the limitations of the previous works as a motivation of the current study.

Why were these kits selected? As there are numerous other kits available as well, the justification for selecting these kits is missing.

In the last paragraph of introduction, authors have mentioned that “However, optimized sample preparation protocols are only available for a handful of species and are missing for many important crop plants”. While they have used only soyabean in this study. So how can they justify that the method optimized in this study for soyabean can work best for other species as well?

Authors have not mentioned how many samples per method they used and how many replicates of each method were used.

In material and methods, authors have measured purity of DNA by Nanodrop and quality and size of DNA using gel electrophoresis, while no such data is given in manuscript. There are numerous

other methods mentioned while no results are given for them.

Figure 5 is of low quality and hard to read, its quality needs to be improved.

**Is the work clearly and accurately presented and does it cite the current literature?**

Yes

**Is the study design appropriate and is the work technically sound?**

No

**Are sufficient details of methods and analysis provided to allow replication by others?**

Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**

Partly

**Are all the source data underlying the results available to ensure full reproducibility?**

No

**Are the conclusions drawn adequately supported by the results?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Epigenetics, Genomics, Rice, Photosynthesis, DNA methylation, Histone posttranslational modifications, RNA-seq

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.**

Reviewer Report 22 February 2023

<https://doi.org/10.5256/f1000research.140976.r162649>

© 2023 Stevens A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Aaron John Stevens** 

Department of Pathology and Molecular Medicine, University of Otago, Wellington, New Zealand

In this article the authors evaluate the performance of different DNA methylation profiling techniques that convert cytosine to uracil and assess their validity for high-throughput sequencing. Two commercial sodium bisulfite (SB) conversion kits developed by Qiagen and Zymo were compared with the NEB EM-seq kit (EM) using genomic DNA extracted from soybean leaf.

The SB method relies on the deamination of non methylated cytosine to uracil using the chemical, sodium bisulfite, whereas EM uses the ten-eleven translocation (TET) enzyme to convert 5-mC to 5-hmC, 5-fC, and eventually to 5-caC, protecting methylated cytosines. The APOBEC3A enzyme is then used to deaminate non-methylated cytosines into uracil. EM conversion has been reported to provide superior results compared to SB treatment, as it offers better data, higher coverage, and more reliable detection of methylation states even with extremely low levels of DNA input. Presumably the main point of difference is that this article investigates plant epigenomes using soybean DNA.

The QIAGEN Conversion kit outperformed the other kits in terms of time requirement and DNA recovery. The NEB EM-seq kit resulted in the lowest DNA recovery, possibly due to more purification steps increasing the chances of DNA loss. However, all three kits yielded quality sequencing data with no major differences in total number of sequencing reads, total coverage, duplication rate, Phred quality score, per-base sequence, and GC content. The results suggest that the cytosine conversion method has little effect on the initial sequence quality, and factors such as time requirement and DNA recovery should be considered when selecting a kit for preparing samples for WGBS analysis of crop plant epigenomes.

**Major concerns:**

- This is a superficial analysis and it does not appear that the authors have considered the use of replicates and instead present the results from a single run of each method. Without replicates there is no way to measure intra kit variation rates, which effectively renders the results meaningless. I have performed hundreds of bisulfite conversions with the zymo kit and will frequently see variation in the conversion ratio, the quality and the yield of DNA between samples from the same input DNA quality and content.
- Have the authors measured returned yield of DNA through the kit against the input amount? This is very important consideration especially when sample is limited.
- The figures and figure caption are not presented in a manner appropriate for scientific publication. For example Figure 3 should be labelled as figure 3A, B, C with a full caption and no legend.
- Authors state several experiments in their methods that they do not include in their results. E.g. gel electrophoresis and that the DNA quantity and quality were measured (neither appear to be stated). However “library recovery” is stated with an inadequate description of what this relates to. Is this the DNA amount measured pre and post conversion? Where is the raw data relating to this?
- Where is the fragment size information prior to sequencing. Should DNA fragment size have been compared prior to the bisulfite conversion also?

**Minor concerns:**

- The first subheading in results/Discussion is a full summary of the section.
- Table 3 should be in methods not in results.

**Is the work clearly and accurately presented and does it cite the current literature?**

Yes

**Is the study design appropriate and is the work technically sound?**

No

**Are sufficient details of methods and analysis provided to allow replication by others?**

Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**

Partly

**Are all the source data underlying the results available to ensure full reproducibility?**

No

**Are the conclusions drawn adequately supported by the results?**

No

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** DNA methylation, genomics, genetics, gene structure, microbiome, rnaseq, bioinformatics

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.**

---

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact [research@f1000.com](mailto:research@f1000.com)

**F1000Research**