Check for updates

Comparative restriction enzyme analysis of methylation (CREAM) reveals methylome variability within a clonal *in vitro* cannabis population

Justin Boissinot[1,2,3,4], Kristian Adamek[5],
Andrew Maxwell Phineas Jones[5], Eric Normandeau[2],
Brian Boyle[2] and Davoud Torkamaneh[1,2,3,4]*

[1]Département de phytologie, Université Laval, Québec, QC, Canada, [2]Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, QC, Canada, [3]Centre de recherche et d'innovation sur les végétaux (CRIV), Université Laval, Québec, QC, Canada, [4]Institut intelligence et données (IID), Université Laval, Québec, QC, Canada, [5]Department of Plant Agriculture, University of Guelph, Guelph, ON, Canada

The primary focus of medicinal cannabis research is to ensure the stability of cannabis lines for consistent administration of chemically uniform products to patients. In recent years, tissue culture has emerged as a valuable technique for genetic preservation and rapid multiplication of cannabis clones. However, there is concern that the physical and chemical conditions of the growing media can induce somaclonal variation, potentially impacting the viability and uniformity of clones. To address this concern, we developed Comparative Restriction Enzyme Analysis of Methylation (CREAM), a novel method to assess DNA methylation patterns and used it to study a population of 78 cannabis clones maintained in tissue culture. Through bioinformatics analysis of the methylome, we successfully detected 2,272 polymorphic methylated regions among the clones. Remarkably, our results demonstrated that DNA methylation patterns were preserved across subcultures within the clonal population, allowing us to distinguish between two subsets of clonal lines used in this study. These findings significantly contribute to our understanding of the epigenetic variability within clonal lines in medicinal cannabis produced through tissue culture techniques. This knowledge is crucial for understanding the effects of tissue culture on DNA methylation and ensuring the consistency and reliability of medicinal cannabis products with therapeutic properties. Additionally, the CREAM method is a fast and affordable technology to get a first glimpse at methylation in a biological system. It offers a valuable tool for studying epigenetic variation in other plant species, thereby facilitating broader applications in plant biotechnology and crop improvement.

1 Introduction

Cannabis (*Cannabis sativa* L.) is one the oldest domesticated plants and has a significant economical and societal impact (Torkamaneh and Jones, 2022). It possesses a long history of human use for fiber, oil, seed, and its medicinal and psychoactive properties (Russo et al., 2008; Bonini et al., 2018). Cannabis is a predominantly dioecious diploid annual herbaceous plant ($2n = 20$) that can accumulate a high quantity of specialized phytocannabinoids within its glandular trichomes (Andre et al., 2016). It is known to produce over 545 potentially bioactive secondary metabolites, including more than 177 cannabinoids, various flavonoids, and a plethora of terpenes (Hanuš and Hod, 2020). Despite a large diversity of metabolites produced, the species is often divided and regulated based on the level of a single psychoactive cannabinoid, Δ^9-tetrahydrocannabinol (THC). In most countries (e.g., Canada, the U.S.A., the E.U.), plants that produce less than 0.3% THC are regulated as hemp, while plants producing 0.3% or more are classified as drug-type. In 2022, the global legal drug-type cannabis market was valued at USD 27.7 billion and is projected to reach USD 82.3 billion by 2027 (Markets and Markets, August 2022[1]). Despite the rapid commercial growth of this crop, its biology remains poorly understood due to its long history of prohibition.

Although cannabis is widely used for medicinal and recreational purposes, there are concerns about the consistency and reproducibility of the derived products. This variation is due to a combination of each plant's genome, as well as the environment in which it is grown, referred to as genotype by environment (GxE) interactions (Booth et al., 2017; Campbell et al., 2019). Within the genomic component, there can be genetic mutations as well as epigenetic differences that can both contribute to differential phenotypic expression. In order to reliably produce consistent extracts, it is critical that they are obtained from genetically stable plants grown under highly controlled conditions. Although cannabis is an outcrossing species with exceptionally high levels of within population variability, clonal propagation methods are relatively easy to use and are optimized to produce uniform populations (Monthony et al., 2021). As a result, in recent years, clonal propagation methods have emerged as the primary method for large-scale production of cannabis. These methods include taking cuttings from selected mother plants (i.e., specific plants with desirable growth characteristics and chemical composition that are maintained in a vegetative stage or in tissue culture for extended period of time), ensuring their proper rooting and growth under controlled conditions, and utilizing specialized cloning media or plant growth regulators to stimulate root development (Adamek et al., 2022). By carefully implementing these methods, growers can achieve consistent and uniform cannabis populations for mass production. However, anecdotal reports indicate that clonal lines tend to decline in quality over time, leading to clones with reduced vigor and lower levels of cannabinoids compared to the original mother plant (Adamek et al., 2022). A recent study documented a

significant amount of intra-plant genetic diversity within a mother plant (Adamek et al., 2022). This diversity could impact the long-term genetic fidelity of clonal lines.

An alternative approach to clonal propagation is micropropagation using plant tissue culture techniques to mass-produce plants in a controlled environment. The compact setup of *in vitro* tissue culture allows for a higher density of plants, minimizing the floor area needed for maintaining mother plants. Importantly, the sterile nature of this technique enables the production of insect-, pathogen-, and virus-free propagules, reducing biotic pressures on the plants (Hesami et al., 2021; Monthony et al., 2021). Typically, it is expected that clones produced *in vitro* using tissue culture techniques will share the same genetics and thus express the same phenotypes. However, somaclonal variations, i.e., phenotypic variations between clones produced in tissue culture that are induced by genetic or epigenetic sources, are extensively reported in the literature (Larkin and Scowcroft, 1981; Bairu et al., 2011). Epigenetic regulation has been identified as a major cause of these variations since it affects the gene expression of seedlings at different growth and developmental stages (Miguel and Marum, 2011; Bednarek and Orłowska, 2020). Epigenetic factors are heritable and potentially reversible modifications who influence gene expression without altering the DNA sequence (Lauria and Rossi, 2011). They include mechanisms such as histone state modifications, noncoding RNAs and DNA methylation, which collectively influence chromatin structure (Lauria and Rossi, 2011). They have been hypothesized to be linked to rejuvenation by regulating plant developmental changes in several plant species (Zhang et al., 2020), including cannabis (Hesami et al., 2023). Among these factors, DNA methylation is widely studied and prevalent in plant genomes. It involves the addition of a methyl group to specific cytosine residues in different contexts (i.e., CG, CHG and CHH, where H stands for A, T or G) (Springer and Schmitz, 2017). Recent studies documented that in a tissue culture setting, modification of DNA methylation patterns in the genome is more common and is associated with changes in DNA sequence, chromosome breaks and activation of transposable elements (TEs) that can influence gene regulation, chromatin inactivation and cell differentiation (Ghosh et al., 2021).

To date, different methods have been developed and applied in different species to profile the methylation landscape across genomes (Yong et al., 2016). These methods vary in DNA input, resolution, genomic region coverage and bioinformatics analysis (Bock, 2012). Currently, selecting a suitable approach requires an in-depth knowledge of these methods. Despite significant decrease in sequencing costs and advances in bioinformatics analysis, whole-genome methylome profiling remains expensive in the context of large-scale studies. Hence, different low-cost approaches such as microarray-based DNA methylation profiling techniques, restriction enzyme-based and reduced representation bisulfite sequencing (RRBS) methods were developed and widely used for detecting methylated regions (Li and Tollefsbol, 2021). Regardless, their application in large populations remains limited. Moreover, despite analyses of the patterns and effects of DNA methylation in plants (Zhang et al., 2018; Ghosh et al., 2021), questions such as the

---

accumulation and location of epimutation sites remain unresolved (Us-Camas et al., 2014; Hazarika et al., 2022).

In the context of cannabis production, determining whether clonal lines derived from tissue culture are uniform is crucial for the consistency and reproducibility of the products. In addition, it is essential for maintaining and preserving germplasm, elite genotypes or parental lines used in breeding programs (Adhikary et al., 2021). Very strict and rigorous quality control and assurance processes as well as the standards related to the safety of cannabis products for medicinal applications require the most precise and regulated production chain (Pusiak et al., 2021; MacCallum et al., 2022). Furthermore, product quality depends on agronomic and environmental factors during plant growth, but also inevitably on the genetic and epigenetic fidelity of the cultivated varieties (Backer et al., 2019). Since micropropagation of uniform clonal lines via tissue culture is fundamental to the cannabis industry, it is thereby critical to study the genetic and epigenetic variations of plants to ensure their long-term stability. The concept of epi/genetic uniformity (or fidelity, stability) can be defined as the absence of variation in the epigenome (epigenetic) and the DNA nucleotide sequences (genetic) within clonal lines.

In this study, we developed a fast and affordable methylotyping method, the Comparative Restriction Enzyme Analysis of Methylation (CREAM), to assess DNA methylation patterns. The CREAM approach, coupled with our bioinformatics pipeline, enabled us to evaluate DNA methylation in a population of 78 cannabis plants representing two clonal lines maintained *in vitro*. This study not only introduces a highly efficient and reliable tool for identifying methylated regions but also provides valuable insights into the methylome uniformity of clonal lines derived from *in vitro* tissue culture.

2 Materials and methods

2.1 Plant material and DNA extraction

The cannabis clonal population used in this study was initiated in March 2019 at the University of Guelph (Ontario, Canada). It was developed using two sister lines (seedlings) derived from a cross of a cultivar exhibiting an *indica*-leaning growth habit and THC and cannabidiol (CBD) levels of approximately 13% and < 1%, respectively (Adamek et al., 2023). Nodal explants from the seedlings were subcultured *in vitro* (about every month) and maintained on DKW Basal Medium with Vitamins (Product ID D2470; Phytotechnology Laboratories, Lenexa, Kansas, USA), 1 mL/L plant preservative mixture (PPM; Plant Cell Technology, Washington, DC, USA), 0.6% agar (w/v) (A360-500; Fisher Scientific, Fair Lawn, New Jersey, USA) and pH adjusted to 5.7 to generate two clonal lines (Page et al., 2021). A total of 78 individual plants considered as clones (from this existing population of two sister clonal lines) of the same chronological age but maintained with different subculture frequencies (i.e., number of subcultures ranging from 6 to 11 every 45 days on average depending on the existing material) were selected (Supplementary Figure 1). The two sister lines that composed the population, named AC-150 and AT-

4, contained 63 and 15 clones that were analyzed, respectively. The maternal material (the first clones) has been destroyed to produce new explants as part of the subculture process in the existing population. DNA samples from the 78 clones were extracted from stem tissue collected at the same time and from the same stem regions (middle of the explant) using a Qiagen DNeasy Plant Mini Kit according to the manufacturer's protocol. The DNA concentration of each sample was quantified using a NanoDrop One spectrophotometer (Thermo Scientific, Waltham, MA) and then diluted to 10 ng/µl. A volume of 10 µl containing 100 ng of DNA was used to prepare each sequencing library for each sample.

2.2 CREAM libraries preparation

Two sequencing libraries were prepared in parallel for each sample with the extracted DNA using the Comparative Restriction Enzyme Analysis of Methylation (CREAM) approach (Figure 1) at the Plateforme d'Analyses Génomiques[2] at the Institut de Biologie Intégrative et des Systèmes (IBIS) of Université Laval, Quebec, Canada. The CREAM method builds on the 3D-GBS approach (de Ronne et al., 2023). Briefly, in both libraries, DNA molecules were cleaved at one end either by the restriction enzymes *Nsi*I (New England Biolabs) or *Pst*I (New England Biolabs), which have distinct restriction sites (5'-ATGCA/T-3' and 5'-CTGCA/G-3', respectively) to anchor DNA fragments at specific locations in the genome. Since they recognize specific sequences of six nucleotides (six-cutters) and each nucleotide has one of the four nucleic bases, these enzymes cut at a theoretical frequency of 4^6 or 4096 base pairs (bp). The combined use of the enzymes reduces this theoretical frequency by half, anchoring the DNA fragments at every 2048 bp and thus covering a larger part of the genome. The actual frequency of restriction sites varies between species and within a same species and is mostly influenced by the percentage of GC bases (Roberts and Murray, 1976; Lindsay and Bird, 1987). The other end of the DNA fragments was cleaved by either *Msp*I (New England Biolabs) or *Hpa*II (New England Biolabs), depending on the library. These enzymes are isoschizomers and share the same restriction site (5'-C/CGG-3') but have different sensitivity to full DNA methylation on the second cytosine of the recognition sequence (at the CpG site), *Msp*I being insensitive and *Hpa*II being sensitive to methylation (Waalwijk and Flavell, 1978). As illustrated in Figure 1A, the variation in sensitivity of the restriction enzymes to DNA methylation results in the generation of fragments of different lengths within the same genomic region. All four possibilities expected from this comparative analysis of the restriction fragments in both libraries, based on the filtering of the fragments with the size selection step, are represented in Figure 1B. After digestion, the sample-specific barcodes and universal adapters were ligated. Since the cohesive ends for the *Msp*I and *Hpa*II restriction sites are identical, the same adapters were used in both libraries. A size selection step using a BluePippin apparatus (Sage Science, Beverley, MA, USA) was performed to capture digested fragments of 200-400 bp. Finally, DNA libraries were amplified by PCR and sequenced with an Illumina NovaSeq 6000
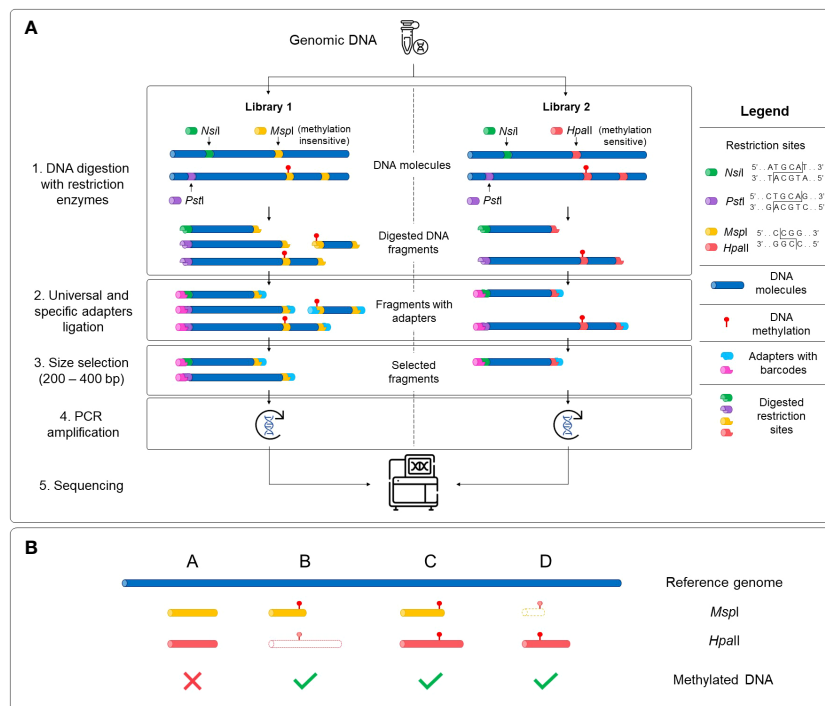
---

2 http://www.ibis.ulaval.ca/en/services-2/genomic-analysis-platform/.

FIGURE 1

Schematic representation of the Comparative Restriction Enzyme Analysis of Methylation (CREAM) approach. **(A)** The same genomic DNA was used as input for the preparation of two libraries. Digestion of the DNA molecules was performed using a set of restriction enzymes sharing a same restriction site but with different sensitivity to DNA methylation (1). Then, universal adapters and sample-specific barcodes were ligated to digested fragments (2). DNA fragments were size selected (3), amplified (4) and sequenced (5). **(B)** The comparative analysis of sequencing data of both libraries leads to four possibilities based on either presence/absence or the length of the DNA fragments. Shared fragments in both libraries with the same length and location in the genome indicate the absence of DNA methylation (possibility A) while differences in the presence or the length of the fragments between the libraries indicate the existence of DNA methylation (possibilities B-D). The DNA fragments with dashed lines and a white center represent absent fragments in the libraries.

System at the Centre d'expertise et de services Génome Québec (Montreal, QC, Canada), generating 204 M and 173 M paired-end reads of 150 bp for the *Msp*I and *Hpa*II libraries, respectively.

2.3 Bioinformatics pipeline

2.3.1 Alignment of reads

The paired-end sequencing reads were demultiplexed using Sabre[3] and trimmed (i.e., removing adapters) with cutadapt (Martin, 2011). Then, they were aligned to the cannabis reference genome (cs10 v2 GenBank Accession No. GCA_900626175.2); (Grassa et al., 2021) with BWA-MEM (Li, 2013). Only reads with a high mapping quality (MAPQ score >= 20) were retained for methylotyping. Genome coverage and depth of coverage were obtained using the bedtools genomecov command (Quinlan and Hall, 2010) and samtools coverage command (Danecek et al., 2021).

2.3.2 Methylotyping

Four possibilities can be expected for the mapped reads (Figure 1B). First, there is no DNA methylation if the fragments are of the same length and mapped to the same location of the

genome in both (*Msp*I and *Hpa*II) libraries (possibility A). The remaining possibilities capture DNA methylation if fragments are found in only one library (possibilities B or D) or if fragments of different lengths are observed (possibility C). To capture these possibilities, a custom pipeline[4] programmed in Python 3[5] was developed and used to determine methylated regions. This pipeline takes the alignments (BAM files) as an input and outputs the methylated and unmethylated positions across the genome. A brief description of the different steps implemented in this pipeline are provided below.

The paired-end reads from the BAM files that are accurately and uniquely mapped to their corresponding pair were used to reconstruct the insert fragment from which they originated. This step ensured that both restriction sites were present in the insert fragments, as DNA methylation can occur on only one side of the restriction fragments. Various quality metrics, including the number of differences between the sequence and the reference (distance), the length of CIGAR strings (indicating insertions or deletions in the sequence), and the number of mismatches in the alignment (alignment score) were extracted from the information generated by BWA-MEM. Subsequently, a table of

---

3   https://github.com/najoshi/sabre.

4   https://github.com/justinboissinot/CREAM.

5   https://www.python.org/.

inserts was generated for each sequencing library, which was further utilized for downstream analysis. Then, from both libraries, a list of loci was extracted based on the inserts. In this context, a locus refers to a genomic region that includes the leftmost and rightmost positions of a set of inserts that either overlap between two libraries or are non-overlapping in either library (Supplementary Figure 2). Loci with overlapping inserts represent regions where the two libraries share at least one nucleotide overlap, including inserts that share the same restriction site or have perfectly overlapping inserts (from their leftmost position to their rightmost position). Loci with non-overlapping inserts represent regions where an insert is present in only one library, capturing regions unique to each library in the loci list. A locus was excluded from the list if less than half of the samples had a coverage of under 20X (inserts) for that specific locus. For each locus, a methylation status was determined based on the aforementioned possibilities (Figure 1B). Methylated and unmethylated positions from the loci were then separated and saved in different files for subsequent analysis.

2.4 Accumulation and distribution of methylated positions

To examine the accumulation of methylated loci in the population (all clones from both sister lines) as the number of subcultures increased, a Kruskal-Wallis test (one-way ANOVA on ranks) was conducted. This test was chosen since the assumptions for conducting an ANOVA were not met in this case. Then, a principal component analysis (PCA) was performed to assess the distribution of methylated positions within the population. The PCA aimed to determine clustering patterns in the samples based on the methylated loci using the R packages *FactoMineR* and *factoextra* (Lê et al., 2008; Kassambara and Mundt, 2016).

The distribution of methylated regions across the cannabis genome was visualized using the *RIdeogram* R package (Hao et al., 2020). This also included the gene density information obtained from the NCBI Gene table for the cs10 reference genome[6] (accessed February 15, 2023). A Spearman's rank correlation coefficient was calculated, using the *cor* command from the R package *stats* (R Core Team, 2022), to assess the monotonic relationship between the gene density and the methylation density across the genome since the data for both variables were skewed towards 0. Other visualizations were generated with *ggplot2* in R (Wickham, 2011).

2.5 Gene ontology analysis

A gene ontology (GO) analysis was performed to identify significant GO terms affected by DNA methylation captured with the CREAM approach. To overcome challenges in matching protein IDs and GO terms with the annotations of the cs10 cannabis reference genome, a combination of the GAWN v0.3.5[7] and go_enrichment[8] pipelines was used. The GAWN pipeline annotated the cs10 reference genome using the available transcriptome from NCBI and found all the methylated loci and the captured loci (unmethylated and methylated) within ± 1 kb of transcripts. The GO enrichment analysis was then performed with the go_enrichment pipeline using the list of methylated loci adjacent to transcripts as the target and considering all captured loci as the background gene set. GO terms with a significant adjusted *p*-value of $p < 0.10$ (Benjamini-Hochberg false discovery rate correction for multiple testing, FDR BH) were kept for further exploration of biological processes.

3 Results

3.1 Development and validation of the CREAM approach

We have developed a low-cost and efficient method for assessing methylome variability at the population level. This method, CREAM, involves digesting DNA samples with restriction enzymes of different sensitivity to DNA methylation. Libraries were developed and sequenced for 78 cannabis clones produced in tissue culture for which no phenotypic data were measured. This has yielded an average of >188 M paired-end reads per sequencing library (Table 1). The demultiplexing, trimming and alignment of the reads to the reference genome led to an average of >2 M paired-end reads per sample.

Out of the initial 78 samples, 10 samples (8 and 2 samples from the AC-150 and AT-4 clonal lines, respectively) were excluded from the analysis as they yielded less than 100,000 reads on average per library. The genome coverage and mean depth of coverage were computed for the remaining samples and compiled for each library (Supplementary Table 1). On average, we captured ~0.4% of the cannabis genome with a mean depth of coverage of ~100X across the captured regions, indicating a sufficient depth to ensure reliable and accurate analysis of the captured regions. As shown in Figure 2, the genome coverage tends to be higher in the *Msp*I library, while the *Hpa*II library exhibits a higher mean depth of coverage. This disparity can be attributed to the fact that the *Hpa*II library does not capture DNA fragments with DNA methylation. Consequently, the *Hpa*II library contains fewer fragments for a comparable sequencing effort compared to the *Msp*I library, covering smaller proportion of the genome with higher coverage.

3.2 Identification of methylated positions

From the mapped paired-end reads, a bioinformatics analytical pipeline was developed to determine the methylated positions captured with the CREAM approach. Briefly, the pipeline

---

6 https://www.ncbi.nlm.nih.gov/data-hub/gene/taxon/3483/.

---

7 https://github.com/enormandeau/gawn.

8 https://github.com/enormandeau/go_enrichment.

TABLE 1 Summary of sequencing libraries statistics.

Feature	Library		
	*Msp*I	*Hpa*II	Overall
Number of raw reads	204 163 492	173 626 876	377 790 368
Average number of reads per sample (after trimming)	2 336 396	1 979 564	2 157 980
Average number of mapped reads per sample	2 324 869 (99.51%)	1 976 345 (99.84%)	2 150 607 (99.66%)

includes quality checks for inserts (composed of a pair of reads) and calls loci that encompass all potentially overlapping inserts within a region (See Supplementary Figure 2 for details). This methylotyping pipeline successfully divided the inserts from the CREAM libraries into the four expected possibilities, resulting in a total of 5,235 loci (See Figure 1B for the four expected categories). Of these confidently called high-quality loci, 3,762 (71.86%) were identified as methylated regions, while 1,473 (28.14%) were identified as unmethylated regions with perfectly overlapping inserts in both libraries (possibility A). The most common type of captured methylated regions (2,949) were loci with an insert only in the *Msp*I library (possibility B), accounting for 56.33% of all captured loci. Loci with an insert only in the *Hpa*II library (possibility D) accounted for approximately 15.53% of all captured loci (813). Finally, a total of 186 loci were excluded from the analysis due to challenges in accurately categorizing them into the predetermined possibilities. These included methylated regions composed of loci with inserts of different length in both libraries (possibility C), which were far less common and loci that exhibited ambiguous

characteristics or lacked clear patterns for classification, rendering them less reliable for further analysis.

3.3 Distribution of methylated positions in the population

The methylotype of each sample for the list of called loci was used to assess the variability in methylated positions across the clonal population. The frequency of the methylated loci revealed four main peaks (Figure 3A). The largest peak consisted of the 1,490 (39.61% of total methylated loci) monomorphic methylated loci, which are DNA methylation positions captured in all 68 remaining samples, indicating a shared methylotype across the population (all clones from both sister clonal lines). On the other hand, among polymorphic loci, 427 (11.35%) were unique to a single sample, indicating specific methylation patterns within individual samples. The two other peaks in the distribution of the number of methylated loci corresponded to the number of clones derived from the two sister lines. Specifically, there were 155 loci unique to the 13 samples in the AT-4 line and 117 loci unique to the 55 samples in the AC-150 line, indicating distinct methylotypes within the sister clonal lines.

In this study, we also investigated whether the number of subcultures that the clonal lines went through in tissue culture had an impact on the number of DNA methylation positions captured. Although a significant methylome variability within clonal lines was observed, we did not find a significant correlation between the number of subcultures and the total number of methylated loci (p-value = 0.19; one-way ANOVA on ranks (Kruskal-Wallis test)) (Figure 3B). This suggests that the variability in DNA methylation within clonal lines was not
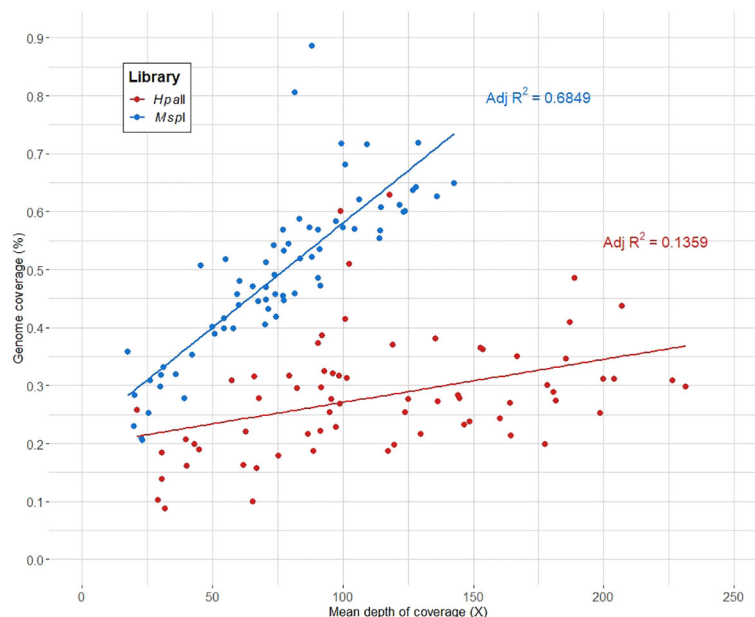

FIGURE 2
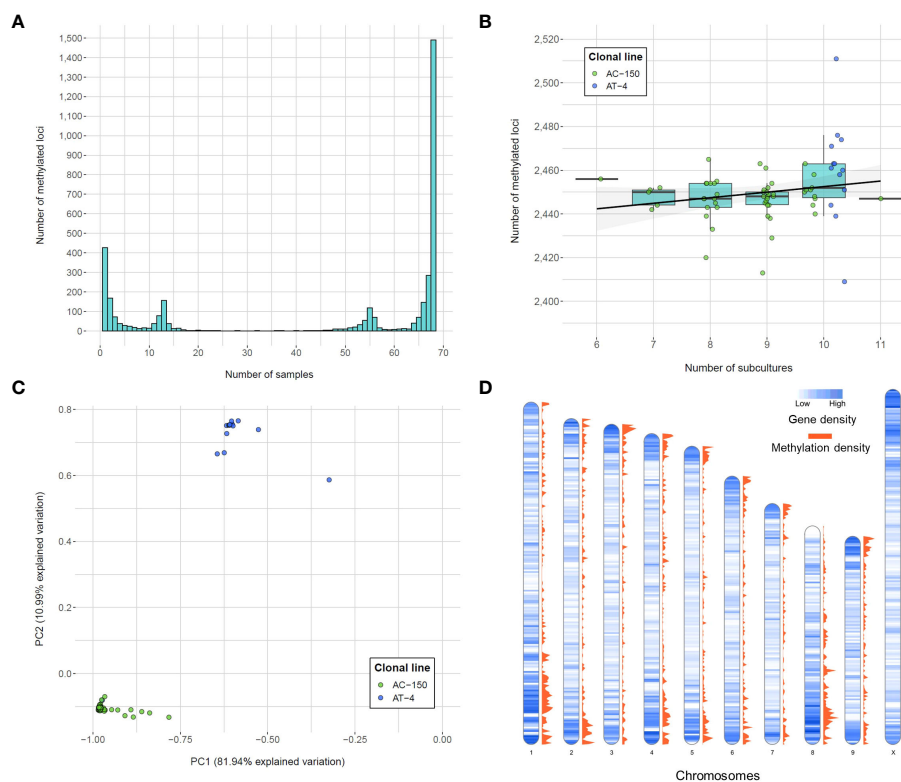Genome coverage (%) per mean depth of coverage (X) for each library.

FIGURE 3
Number of methylated loci called from the CREAM libraries. **(A)** Frequency of methylated loci within the population. **(B)** Number of methylated loci per number of subcultures for both clonal lines (AC-150 line in green and AT-4 line in blue). **(C)** Principal component analysis (PCA) of methylated loci called from the CREAM libraries. Both clonal lines derived from two sister lines are clustered into two different groups. **(D)** Density of genes and methylated loci on the chromosomes of the cs10 cannabis reference genome in bins of 500 kb.

influenced by the number of subcultures the plants went through in this study. Finally, the principal component analysis (PCA) performed on the methylated loci (Figure 3C) revealed that the two principal components explained a significant portion of the variation, accounting for 81.94% and 10.99% of the total variation, respectively. The samples displayed clear clustering patterns, with two distinct groups corresponding to the clones derived from the two sister lines. Within each cluster, the variability was relatively limited, although some samples showed variation compared to overall trend of their respective group.

3.4 Distribution of the methylated positions in the genome

We examined the distribution of the methylated positions across the genome to understand the patterns of DNA methylation. Precisely, we calculated the gene density, which represents the number of genes present in each bin of 500 kb in the cannabis genome, along with the methylation density, which indicates the number of captured methylated positions within the same 500 kb bin (Figure 3D). We then compared the gene density and methylation density to understand the relationship between DNA methylation and gene distribution in different genomic regions. The density of captured methylated loci across the genome showed a strong positive

correlation with the density of genes across the same genomic bins (Spearman's rank correlation coefficient = 0.6473), suggesting the positive monotonic relationship between the two variables. We also observed that regions of high genic density were also enriched in captured methylated loci. However, the observed correlation between gene density and methylation density could be influenced by the approach used, which tends to capture loci in genic regions (Supplementary Figure 3). Finally, the gene ontology (GO) analysis was performed to identify genes potentially affected by the DNA methylation patterns captured within a range of ± 1 kb of gene transcripts (Table 2). The analysis utilized all the loci captured by the CREAM approach as a background set for comparison. As a result, 11 GO terms were found to be significant. Most importantly, 9 out of the 11 significant biological process GO terms were related to metabolic processes, indicating a potential influence of DNA methylation on important biochemical pathways.

4 Discussion

In cannabis and other plant species produced through micropropagation, it is crucial to understand the underlying factors contributing to somaclonal variation. While there has been extensive research on the influence of media culture conditions and genetic variation (Sato et al., 2011; Zhang et al., 2014; Krishna et al., 2016),

TABLE 2   List of enriched GO terms that are influenced by DNA methylation in cannabis.

GO term	Name of biological process	p-value (FDR BH)
GO:0006139	Nucleobase-containing compound metabolic process	0.00001
GO:0046483	Heterocycle metabolic process	0.00001
GO:0090304	Nucleic acid metabolic process	0.00003
GO:0006725	Cellular aromatic compound metabolic process	0.0001
GO:1901360	Organic cyclic compound metabolic process	0.0002
GO:0034641	Cellular nitrogen compound metabolic process	0.0003
GO:0016070	RNA metabolic process	0.009
GO:0006807	Nitrogen compound metabolic process	0.011
GO:0016071	mRNA metabolic process	0.045
GO:0009987	Cellular process	0.084
GO:0006325	Chromatin organization	0.084

epigenetic factors have recently emerged as noteworthy factors that could account for phenotypic variability that cannot be explained by genetic mutations in micropropagated plants. Of particular interest is the examination of DNA methylation patterns (Jaligot et al., 2000; Matthes et al., 2001; Li et al., 2012; Bobadilla Landey et al., 2015; Ong-Abdullah et al., 2015; Han et al., 2018; Borges et al., 2021; Wibowo et al., 2022). Despite an increasing interest in cannabis as a valuable crop, there is a notable gap in research regarding DNA methylation and its influence on the cannabis genome, particularly in the context of tissue culture. This knowledge gap hinders the development of effective strategies to ensure reproducibility and efficiency in tissue culture practices, which are essential for various production systems. Therefore, the primary objective of this study was to develop a novel, low-cost, and high-throughput approach (i.e., CREAM) to detect variations in the methylome of *C. sativa* clones derived from *in vitro* tissue culture. By investigating DNA methylation patterns, we aimed to shed light on epigenetic factors contributing to somaclonal variation in cannabis and provide insights into the potential regulatory role of DNA methylation in shaping the phenotypic diversity observed among clones.

4.1 CREAM: a new efficient and cost-effective methylotyping method

The CREAM method was successful in identifying significant variation in methylotypes among cannabis clones, specifically distinguishing the two subpopulations derived from two sister lines.

This success can be attributed to the generation of high-quality sequencing fragments, which facilitated the accurate identification of methylated and unmethylated loci using our newly developed methylotyping pipeline. Importantly, the CREAM method offers a cost-effective solution compared to the main approaches based on RRBS, like epiGBS (Van Gurp et al., 2016) and bsRADseq (Trucchi et al., 2016), with an estimated sequencing cost per sample of around 30 $, representing a 30-80% decrease in costs (Werner et al., 2020). Additionally, the CREAM approach avoided the extensive DNA damage typically associated with the bisulfite conversion in other techniques (Tanaka and Okamoto, 2007), preserving DNA integrity. Other large-scale methylation detection approaches have been developed to investigate methylation patterns on a large scale, such as Nanopore sequencing (Tourancheau et al., 2021) and whole-genome bisulfite sequencing (Urich et al., 2015). These established methods have demonstrated enhanced success in identifying methylation patterns (Ni et al., 2021). Additionally, ongoing efforts to enhance methylation calling (Ni et al., 2021) have further improved detection rates. However, while these methods may offer a thorough and insightful evaluation of methylated regions across the genome, they are still generally considered expensive to conduct studies at a population level (Suzuki et al., 2018; Kong et al., 2023).

The CREAM method achieved a satisfactory average genome coverage of ~0.4% and a mean depth of coverage of around 100X. In comparison, well-regarded RRBS methods such as epiGBS (Van Gurp et al., 2016) and its modified version (Werner et al., 2020) covered 0.37% of the 135 Mb of the *A. thaliana* genome and 0.28% of the 246 Mb almond genome, respectively. When examining the genome coverage and mean depth of coverage specific to each library (Figure 2), it becomes evident that the genome coverage tends to be higher in the *Msp*I library compared to the *Hpa*II library. This observation is consistent with the higher proportion of loci with fragments found only in the *Msp*I library (category B) among the captured regions. Therefore, the differences in genome coverage and the number of loci found between the two libraries can be attributed to the variations in the efficiency of the two restriction enzymes (methylation sensitive and methylation insensitive) used in the CREAM method.

The CREAM method builds on methylation-sensitive amplified polymorphism (MSAP) approaches coupled with sequencing technologies, which aim to provide a direct genomic localization of differentially cleaved DNA fragments using methylation-sensitive restriction enzymes. Typically, DNA fragments are cleaved at one end by the *Eco*RI enzyme and at the other end by one of the two isoschizomers *Msp*I or *Hpa*II (restriction enzymes sharing the same restriction site, 5'-CCGG-3'). The presence and absence of *Eco*RI/*Msp*I (unmethylated) and *Eco*RI/*Hpa*II (methylated) fragments allows methylation patterns to be determined in a comparative manner (Fulneček and Kovařík, 2014), and their sequencing allows for their genomic localization. Such methods, like MSAP-Seq (Chwialkowska et al., 2017) and others (Baránek et al., 2016) have achieved the direct analysis of DNA methylation in plants using only *Eco*RI as an anchoring enzyme in combination with *Msp*I and *Hpa*II. Using a permissively low coverage of 2 reads per sample, the MSAP-Seq was able to capture a total of 190,000 CCGG sites for DNA methylation analysis in barley, which has a genome of about

5.3 Gb (around 6 times the size of the cannabis genome) (Chwialkowska et al., 2017). The CREAM method focuses on higher quality CCGG sites, with a coverage of at least 20, to avoid artefacts in captured methylated sites. It also utilizes two anchoring enzymes (NsiI and PstI) to capture more diverse restriction fragments, therefore providing a better overview of methylation sites across the genome. While it provides interesting elements to the family of MSAP methods coupled with sequencing technologies, the CREAM approach also only focuses on CCGG sequences recognized by restriction enzymes, thus ignoring DNA methylation in CHG and CHH contexts, notably involved in the regulation of genes and transposons (Agius et al., 2023).

The choice of the restriction enzymes in the CREAM method has a direct impact on the number of loci captured and their distribution across the genome. In our findings, we observed a notable enrichment of fragments within genic regions. To explain the abundance of captured fragments, we propose the influence of GC content as a contributing factor. In the case of the cannabis reference genome (cs10), it exhibits an overall GC content of 33%, whereas the GC content differs among the restriction sites of the restriction enzymes. Specifically, the GC content of the NsiI and PstI restriction sites is 66% and 33%, respectively, while the MspI and HpaII restriction sites have a GC content of 100%. This indicates that the distribution of restriction fragments is not uniform across the genome and tends to be more concentrated in regions that are relatively richer in GC than AT. In plants, these GC-rich regions are predominantly associated with genic regions (Serres-Giardi et al., 2012; Glémin et al., 2014). This supports the observed correlation between the gene density and methylation density in the data (Figure 3D). Therefore, the CREAM approach might result in an overrepresentation of genic regions due to their higher accessibility to the restriction enzymes.

While our approach offers several advantages, it is important to acknowledge its limitations. First, its resolution is currently limited. It provides a binary methylation status (either 0 or 1) for a given cytosine in a locus, rather than providing a methylation quantitative score at a base level. Although this binary representation still provides valuable insights, a finer resolution would be desirable to gain a more comprehensive understanding of the methylation landscape across the genome. Second, we encountered challenges in classifying 186 captured loci (3.43% of all loci) into the expected categories (A, B, C or D). The comparative analysis of the libraries was challenging for these specific restriction fragments because of the complexity and diversity of their alignment. Lastly, it is important to note that our approach did not operate at a single-cell level. As a result, the methylation patterns obtained from a single sample represent a mixture of methylation patterns from various cells, introducing some level of heterogeneity. However, these limitations might be overshadowed by the compelling low-cost and high-throughput features of the approach.

4.2 Methylation variation in cannabis clones

Upon analyzing the DNA methylation frequency in the population, we observed contrasting patterns between the two subsets of clonal lines. These patterns can be used as methylation tags or methylation fingerprints, allowing us to clearly differentiate individual clones within their respective clonal line. Notably, the methylation patterns exhibited by each clone strongly reflect their clonal lineage, suggesting a unique and identifiable DNA methylation profile for each clonal line. This is consistent with various studies of methylation patterns in tissue culture. For example, in maize, stable changes in differentially methylated regions (DMRs) were observed between independently regenerated clonal lines, highlighting potential "epigenetic footprints" unique to certain lines and maintained over several generations (Stelpflug et al., 2014). A further study has shown consistent (monomorphic) and rare (polymorphic) DMRs maintained across generations that distinguished tissue culture-derived plants, providing more evidence for specific DNA methylation patterns acting as "tags" (Han et al., 2018). It is likely that these patterns originated from the sister seeds that were used to initiate clonal lines and were preserved to some extent throughout multiple generations of subculturing. In sexual reproduction, such as seed production, various mechanisms exist to limit defective and potentially problematic epigenetic variations from one generation to the next, including resetting of DNA methylation state (Quadrana and Colot, 2016). However, clonal propagation, being asexual, lacks these reset processes and is expected to affect DNA methylation stability across the genome (Ibañez and Quadrana, 2023), leading to potential slight variations in methylation patterns.

Different DNA methylation dynamics have been observed in the vegetative state. For instance, high levels of CHG methylation in Arabidopsis thaliana shoot apical meristematic stem cells (Gutzat et al., 2020) and low levels of CHH methylation in species with extensive clonal propagation histories (Niederhuth et al., 2016) have been reported. However, trends regarding CG methylation levels in tissue culture are less conclusive and differ among species, with increasing levels reported in gentian (Fiuk et al., 2010), bush lily (Wang et al., 2012), banana (Peraza-Echeverria et al., 2001) and tomato (Smulders et al., 1995), decreasing levels observed in triticale (Bednarek et al., 2017), barley (Li et al., 2007), grapevine (Baránek et al., 2010) and Freesia (Gao et al., 2010), and no significant changes detected in pea (Smýkal et al., 2007) or apples (Li et al., 2002). Since the cannabis methylome is poorly understood, our results represent an encouraging first report of methylation patterns for this economically important crop in a tissue culture context.

Tissue-specific differences in DNA methylation can also be significant (Lloyd and Lister, 2022). Cells from stem tissues are generally less studied and methylation patterns are less decisive, making comparisons between studies and species more intricate. While no methylome studies have been done in cannabis, it is still interesting to present the results of this current study in the context of the closest work achieved in terms of species of interest, DNA methylation capture approach and stem tissue studied in a tissue culture propagation system. In a study conducted in hops (Humulus lupulus L.), a close relative to C. sativa (Kovalchuk et al., 2020), micropropagated stem tissues (branches) were tested for methylation changes (Peredo et al., 2009). The study found that most the methylated loci (56.34%) were monomorphic, meaning they were shared by all 80 clones obtained from two cultivars, while 13.24% of the variation were unique to individual samples (referred

to as singletons). These results are consistent with the distribution of methylated loci observed in the present study (Figure 3A), with 41.22% and 11.81% of the total loci being monomorphic and unique to a single sample, respectively. The difference in proportions between the methylation patterns observed in our study and those reported in the MSAP method used in the hops study (Peredo et al., 2009) could be attributed to the higher genomic resolution of our approach and the variability between the species. Further methylome studies in cannabis would also benefit from tracking methylation patterns in several tissues, such as female flowers that accumulate many compounds of interest.

The lack of significant differences in the number of methylated loci with the number of subcultures may be attributed to the dynamic coordination of processes involved in establishing, maintaining, and removing specific DNA methylation states (Zhang et al., 2018). It has been reported that while the total number of methylated positions may not change significantly, their specific identity is likely to vary (Zhang et al., 2018). This suggests that the stability of methylation patterns is established in the earlier stages of production. The stability of DNA methylation patterns during subculturing has been observed in other plant species as well. For example, in garlic, methylation patterns were found to stabilize after 6 months of micropropagation (Gimenez et al., 2016). However, it is important to note that the trends in DNA methylation variations are highly species-dependant. Some studies have reported increases in methylation levels during subculturing (Rival et al., 2013; Fraga et al., 2016), while others have observed decreases (Xu et al., 2004; Huang et al., 2012; Machczyńska et al., 2014) over different subculturing durations. There have also been cases where methylation levels initially decrease in early generations and then recover over longer periods of time, as shown in sweet orange after 30 years (Wang et al., 2022). Considering that our study is the first investigation of the cannabis methylome, it is plausible that methylation patterns stabilized in our clonal population over several subcultures. All clones were acclimatized to tissue culture and could have reached a steady state by the 6th to 10th subculture. Further research is needed to gain a better understanding of the dynamics and stability of DNA methylation patterns in cannabis and how they may influence the phenotypic characteristics of clonal plants. A focus should be made on the induction phase in tissue culture, where plants are transitioning from regular growth conditions to *in vitro* conditions, which could affect methylation patterns.

4.3 Distribution of methylated loci in the cannabis genome

Upon examining the distribution of DNA methylation in the cannabis genome, we observed a correlation between methylation density and gene density along the chromosomes. As mentioned earlier, this suggests that the captured loci in this study may be biased towards genic regions due to their GC-rich content. To a lesser extent, non-genic regions were also captured with the CREAM approach. These genomic regions rich in transposable elements and repetitive DNA sequences have also been associated

with higher DNA methylation rates in *A. thaliana*, both in heterochromatin and euchromatin (Zhang et al., 2006). It is therefore expected to observe methylated loci in these regions. With the current limited knowledge of the cannabis genome, loci found near genes still represent the most interesting regions to study for their immediate potential functional impact on the plants biological processes.

In this study, the GO analysis within a range of ± 1 kb around gene transcripts resulted in 11 significant GO terms. The effect of gene body methylation (gbM) on gene expression and phenotypic changes is still a subject of debate and not fully understood. However, the association between gbM and gene expression is recognized as important, and it represents a potential source of variation that can be subject to natural selection (Muyle et al., 2022), highlighting the importance of the significant GO terms found in this study. Of these 11 GO terms, 9 are related to metabolic processes, which encompass a range of biochemical reactions involved in the synthesis, breakdown, and transformation of various molecules within the plant. This suggests that DNA methylation may play a regulatory role in modulating the expression and activity of genes involved in key metabolic pathways in cannabis. To gain a deeper understanding of the cannabis methylome, it would be crucial to validate the biological effects of DNA methylation on the transcriptional activity of these genes. Such validation experiments would provide important insights into the functional implications of DNA methylation in cannabis and its potential impact on metabolic processes.

In conclusion, the present study addresses a critical gap regarding somaclonal variation and epigenetic factors in *C. sativa*. To the best of our knowledge, this is the first study to investigate the cannabis methylome and the first of its kind in the field of cannabis tissue culture. The primary objective of this study was to develop a novel and cost-effective approach to detect methylation variation in *C. sativa* clones derived from tissue culture. The results revealed significant variation in methylotypes among the population of cannabis clones, indicating the presence of methylation footprints between sister clonal lines and offering valuable insights into the epigenetic landscape of this important crop. Importantly, this approach overcomes the cost and technical challenges associated with existing methods and enables the high-throughput analysis of methylome at a population level. This knowledge has implications for crop improvement and the development of sustainable production systems in the cannabis industry. By highlighting the dynamics of methylome and its correlation with gene density, this study advocates for further investigations on the regulatory role of DNA methylation in key metabolic pathways of cannabis. Functional validation of the observed methylation patterns and their impact on gene expression will undoubtedly contribute to a more comprehensive understanding of the cannabis methylome.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

JB: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. KA: Data curation, Investigation, Methodology, Writing – review & editing. AJ: Data curation, Investigation, Methodology, Writing – review & editing. EN: Conceptualization, Investigation, Methodology, Validation, Writing – review & editing. BB: Conceptualization, Data curation, Investigation, Methodology, Resources, Writing – review & editing. DT: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

Funding

Acknowledgments

Conflict of interest

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

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1381154/full#supplementary-material

References

Adamek, K., Grainger, C., Jones, A. M. P., and Torkamaneh, D. (2023). Genotyping-by-sequencing (GBS) reveals greater somatic mutations than simple sequence repeats (SSRs) in micropropagated cannabis plants. *Vitro Cell. Dev. Biol. - Plant* 59, 757–766. doi: 10.1007/s11627-023-10377-x

Adamek, K., Jones, A. M. P., and Torkamaneh, D. (2022). Accumulation of somatic mutations leads to genetic mosaicism in cannabis. *Plant Genome* 15. doi: 10.1002/tpg2.20169

Adhikary, D., Kulkarni, M., El-Mezawy, A., Mobini, S., Elhiti, M., Gjuric, R., et al. (2021). Medical cannabis and industrial hemp tissue culture: present status and future potential. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.627240

Agius, D. R., Kapazoglou, A., Avramidou, E., Baranek, M., Carneros, E., Caro, E., et al. (2023). Exploring the crop epigenome: a comparison of DNA methylation profiling techniques. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1181039

Andre, C. M., Hausman, J.-F., and Guerriero, G. (2016). *Cannabis sativa*: the plant of the thousand and one molecules. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00019

Backer, R., Schwinghamer, T., Rosenbaum, P., McCarty, V., Eichhorn Bilodeau, S., Lyu, D., et al. (2019). Closing the yield gap for cannabis: A meta-analysis of factors determining cannabis yield. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00495

Bairu, M. W., Aremu, A. O., and Van Staden, J. (2011). Somaclonal variation in plants: causes and detection methods. *Plant Growth Regul.* 63, 147–173. doi: 10.1007/s10725-010-9554-x

Baránek, M., Čechová, J., Kovacs, T., Eichmeier, A., Wang, S., Raddová, J., et al. (2016). Use of combined MSAP and NGS techniques to identify differentially methylated regions in somaclones: A case study of two stable somatic wheat mutants. *PloS One* 11, e0165749. doi: 10.1371/journal.pone.0165749

Baránek, M., Křižan, B., Ondrušíková, E., and Pidra, M. (2010). DNA-methylation changes in grapevine somaclones following *in vitro* culture and thermotherapy. *Plant Cell Tissue Organ Cult. PCTOC* 101, 11–22. doi: 10.1007/s11240-009-9656-1

Bednarek, P. T., and Orłowska, R. (2020). Plant tissue culture environment as a switch-key of (epi)genetic changes. *Plant Cell Tissue Organ Cult. PCTOC* 140, 245–257. doi: 10.1007/s11240-019-01724-1

Bednarek, P. T., Orłowska, R., and Niedziela, A. (2017). A relative quantitative Methylation-Sensitive Amplified Polymorphism (MSAP) method for the analysis of abiotic stress. *BMC Plant Biol.* 17, 79. doi: 10.1186/s12870-017-1028-0

Bobadilla Landey, R., Cenci, A., Guyot, R., Bertrand, B., Georget, F., Dechamp, E., et al. (2015). Assessment of genetic and epigenetic changes during cell culture ageing and relations with somaclonal variation in *Coffea arabica*. *Plant Cell Tissue Organ Cult. PCTOC* 122, 517–531. doi: 10.1007/s11240-015-0772-9

Bock, C. (2012). Analysing and interpreting DNA methylation data. *Nat. Rev. Genet.* 13, 705–719. doi: 10.1038/nrg3273

Boissinot, J., Adamek, K., Jones, A. M. P., Normandeau, E., Boyle, B., and Torkamaneh, D. (2023). Comparative restriction enzyme analysis of methylation (CREAM) reveals methylome variability within a clonal *in vitro* cannabis population. *bioRxiv*. doi: 10.1101/2023.08.18.552785

Bonini, S. A., Premoli, M., Tambaro, S., Kumar, A., Maccarinelli, G., Memo, M., et al. (2018). *Cannabis sativa*: A comprehensive ethnopharmacological review of a medicinal plant with a long history. *J. Ethnopharmacol* 227, 300–315. doi: 10.1016/j.jep.2018.09.004

Booth, J. K., Page, J. E., and Bohlmann, J. (2017). Terpene synthases from *Cannabis sativa*. *PloS One* 12, e0173911. doi: 10.1371/journal.pone.0173911

Borges, F., Donoghue, M. T. A., LeBlanc, C., Wear, E. E., Tanurdžić, M., Berube, B., et al. (2021). Loss of small-RNA-directed DNA methylation in the plant cell cycle promotes germline reprogramming and somaclonal variation. *Curr. Biol.* 31, 591–600.e4. doi: 10.1016/j.cub.2020.10.098

Campbell, B. J., Berrada, A. F., Hudalla, C., Amaducci, S., and McKay, J. K. (2019). Genotype × Environment interactions of industrial hemp cultivars highlight diverse responses to environmental factors. *Agrosystems Geosci. Environ.* 2, 1–11. doi: 10.2134/age2018.11.0057

Chwialkowska, K., Korotko, U., Kosinska, J., Szarejko, I., and Kwasniewski, M. (2017). Methylation sensitive amplification polymorphism sequencing (MSAP-seq)—A method for high-throughput analysis of differentially methylated CCGG sites in plants with large genomes. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.02056

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* 10, giab008. doi: 10.1093/gigascience/giab008

de Ronne, M., Légaré, G., Belzile, F., Boyle, B., and Torkamaneh, D. (2023). 3D-GBS: a universal genotyping-by-sequencing approach for genomic selection and other high-throughput low-cost applications in species with small to medium-sized genomes. *Plant Methods* 19, 13. doi: 10.1186/s13007-023-00990-7

Fiuk, A., Bednarek, P. T., and Rybczyński, J. J. (2010). Flow Cytometry, HPLC-RP, and metAFLP Analyses to Assess Genetic Variability in Somatic Embryo-Derived Plantlets of *Gentiana pannonica* Scop. *Plant Mol. Biol. Rep.* 28, 413–420. doi: 10.1007/s11105-009-0167-3

Fraga, H. P. F., Vieira, L. N., Heringer, A. S., Puttkammer, C. C., Silveira, V., and Guerra, M. P. (2016). DNA methylation and proteome profiles of *Araucaria angustifolia* (Bertol.) Kuntze embryogenic cultures as affected by plant growth regulators supplementation. *Plant Cell Tissue Organ Cult. PCTOC* 125, 353–374. doi: 10.1007/s11240-016-0956-y

Fulneček, J., and Kovařík, A. (2014). How to interpret Methylation Sensitive Amplified Polymorphism (MSAP) profiles? *BMC Genet.* 15, 2. doi: 10.1186/1471-2156-15-2

Gao, X., Yang, D., Cao, D., Ao, M., Sui, X., Wang, Q., et al. (2010). *In vitro* microprogagation of *freesia hybrida* and the assessment of genetic and epigenetic stability in regenerated plantlets. *J. Plant Growth Regul.* 29, 257–267. doi: 10.1007/s00344-009-9133-4

Ghosh, A., Igamberdiev, A. U., and Debnath, S. C. (2021). Tissue culture-induced DNA methylation in crop plants: a review. *Mol. Biol. Rep.* 48, 823–841. doi: 10.1007/s11033-020-06062-6

Gimenez, M. D., Yañez-Santos, A. M., Paz, R. C., Quiroga, M. P., Marfil, C. F., Conci, V. C., et al. (2016). Assessment of genetic and epigenetic changes in virus-free garlic (*Allium sativum* L.) plants obtained by meristem culture followed by *in vitro* propagation. *Plant Cell Rep.* 35, 129–141. doi: 10.1007/s00299-015-1874-x

Glémin, S., Clément, Y., David, J., and Ressayre, A. (2014). GC content evolution in coding regions of angiosperm genomes: a unifying hypothesis. *Trends Genet.* 30, 263–270. doi: 10.1016/j.tig.2014.05.002

Grassa, C. J., Weiblen, G. D., Wenger, J. P., Dabney, C., Poplawski, S. G., Timothy Motley, S., et al. (2021). A new *Cannabis* genome assembly associates elevated cannabidiol (CBD) with hemp introgressed into marijuana. *New Phytol.* 230, 1665–1679. doi: 10.1111/nph.17243

Gutzat, R, Rembart, K, Nussbaumer, T, Hofmann, F, Pisupati, R, Bradamante, G., et al. (2020). *Arabidopsis* shoot stem cells display dynamic transcription and DNA methylation patterns. *EMBO J.* 39, e103667. doi: 10.15252/embj.2019103667

Han, Z., Crisp, P. A., Stelpflug, S., Kaeppler, S. M., Li, Q., and Springer, N. M. (2018). Heritable epigenomic changes to the maize methylome resulting from tissue culture. *Genetics* 209, 983–995. doi: 10.1534/genetics.118.300987

Hanuš, L. O., and Hod, Y. (2020). Terpenes/terpenoids in *cannabis*: are they important? *Med. Cannabis Cannabinoids* 3, 25–60. doi: 10.1159/000509733

Hao, Z., Lv, D., Ge, Y., Shi, J., Weijers, D., Yu, G., et al. (2020). RIdeogram : drawing SVG graphics to visualize and map genome-wide data on the idiograms. *PeerJ Comput. Sci.* 6, e251. doi: 10.7717/peerj-cs.251

Hazarika, R. R., Serra, M., Zhang, Z., Zhang, Y., Schmitz, R. J., and Johannes, F. (2022). Molecular properties of epimutation hotspots. *Nat. Plants* 8, 146–156. doi: 10.1038/s41477-021-01086-7

Hesami, M., Adamek, K., Pepe, M., and Jones, A. M. P. (2023). Effect of explant source on phenotypic changes of *in vitro* grown cannabis plantlets over multiple subcultures. *Biology* 12, 443. doi: 10.3390/biology12030443

Hesami, M., Baiton, A., Alizadeh, M., Pepe, M., Torkamaneh, D., and Jones, A. M. P. (2021). Advances and perspectives in tissue culture and genetic engineering of cannabis. *Int. J. Mol. Sci.* 22, 5671. doi: 10.3390/ijms22115671

Huang, H., Han, S. S., Wang, Y., Zhang, X. Z., and Han, Z. H. (2012). Variations in leaf morphology and DNA methylation following *in vitro* culture of *Malus xiaojinensis*. *Plant Cell Tissue Organ Cult. PCTOC* 111, 153–161. doi: 10.1007/s11240-012-0179-9

Ibañez, V. N., and Quadrana, L. (2023). Shaping inheritance: how distinct reproductive strategies influence DNA methylation memory in plants. *Curr. Opin. Genet. Dev.* 78, 102018. doi: 10.1016/j.gde.2022.102018

Jaligot, E., Rival, A., Beulé, T., Dussert, S., and Verdeil, J.-L. (2000). Somaclonal variation in oil palm (*Elaeis guineensis* Jacq.): the DNA methylation hypothesis. *Plant Cell Rep.* 19, 684–690. doi: 10.1007/s002999900177

Kassambara, A., and Mundt, F. (2016). *Package 'factoextra' - Extract and Visualize the Results of Multivariate Data Analyses*.

Kong, Y., Mead, E. A., and Fang, G. (2023). Navigating the pitfalls of mapping DNA and RNA modifications. *Nat. Rev. Genet.* 24, 363–381. doi: 10.1038/s41576-022-00559-5

Kovalchuk, I., Pellino, M., Rigault, P., van Velzen, R., Ebersbach, J., Ashnest, J. R., et al. (2020). The genomics of *cannabis* and its close relatives. *Annu. Rev. Plant Biol.* 71, 713–739. doi: 10.1146/annurev-arplant-081519-040203

Krishna, H., Alizadeh, M., Singh, D., Singh, U., Chauhan, N., Eftekhari, M., et al. (2016). Somaclonal variations and their applications in horticultural crops improvement. *3 Biotech.* 6, 54. doi: 10.1007/s13205-016-0389-7

Larkin, P. J., and Scowcroft, W. R. (1981). Somaclonal variation — a novel source of variability from cell cultures for plant improvement. *Theor. Appl. Genet.* 60, 197–214. doi: 10.1007/BF02342540

Lauria, M., and Rossi, V. (2011). Epigenetic control of gene regulation in plants. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* 1809, 369–378. doi: 10.1016/j.bbagrm.2011.03.002

Lê, S., Josse, J., and Husson, F. (2008). FactoMineR : an *R* package for multivariate analysis. *J. Stat. Softw* 25. doi: 10.18637/jss.v025.i01

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. doi: 10.48550/arXiv.1303.3997

Li, S., and Tollefsbol, T. O. (2021). DNA methylation methods: Global DNA methylation and methylomic analyses. *Methods* 187, 28–43. doi: 10.1016/j.ymeth.2020.10.002

Li, X., Xu, M., and Korban, S. S. (2002). DNA methylation profiles differ between field- and *in vitro*-grown leaves of apple. *J. Plant Physiol.* 159, 1229–1234. doi: 10.1078/0176-1617-00899

Li, X., Yu, X., Wang, N., Feng, Q., Dong, Z., Liu, L., et al. (2007). Genetic and epigenetic instabilities induced by tissue culture in wild barley (*Hordeum brevisubulatum* (Trin.) Link). *Plant Cell Tissue Organ Cult* 90, 153–168. doi: 10.1007/s11240-007-9224-5

Li, H., Zhao, X., Dai, H., Wu, W., Mao, W., and Zhang, Z. (2012). Tissue culture responsive microRNAs in strawberry. *Plant Mol. Biol. Rep.* 30, 1047–1054. doi: 10.1007/s11105-011-0406-2

Lindsay, S., and Bird, A. P. (1987). Use of restriction enzymes to detect potential gene sequences in mammalian DNA. *Nature* 327, 336–338. doi: 10.1038/327336a0

Lloyd, J. P. B., and Lister, R. (2022). Epigenome plasticity in plants. *Nat. Rev. Genet.* 23, 55–68. doi: 10.1038/s41576-021-00407-y

MacCallum, C. A., Lo, L. A., Pistawka, C. A., and Boivin, M. (2022). A clinical framework for evaluating cannabis product quality and safety. *Cannabis Cannabinoid Res.* doi: 10.1089/can.2021.0137

Machczyńska, J., Orłowska, R., Mańkowski, D. R., Zimny, J., and Bednarek, P. T. (2014). DNA methylation changes in triticale due to *in vitro* culture plant regeneration and consecutive reproduction. *Plant Cell Tissue Organ Cult. PCTOC* 119, 289–299. doi: 10.1007/s11240-014-0533-1

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, 10–12. doi: 10.14806/ej.17.1.200

Matthes, M., Singh, R., Cheah, S.-C., and Karp, A. (2001). Variation in oil palm (*Elaeis guineensis* Jacq.) tissue culture-derived regenerants revealed by AFLPs with methylation-sensitive enzymes. *Theor. Appl. Genet.* 102, 971–979. doi: 10.1007/s001220000491

Miguel, C., and Marum, L. (2011). An epigenetic view of plant cells cultured *in vitro*: somaclonal variation and beyond. *J. Exp. Bot.* 62, 3713–3725. doi: 10.1093/jxb/err155

Monthony, A. S., Page, S. R., Hesami, M., and Jones, A. M. P. (2021). The past, present and future of *Cannabis sativa* tissue culture *Plants* 10, 185. doi: 10.3390/plants10010185

Muyle, A. M., Seymour, D. K., Lv, Y., Huettel, B., and Gaut, B. S. (2022). Gene body methylation in plants: mechanisms, functions, and important implications for understanding evolutionary processes. *Genome Biol. Evol.* 14, evac038. doi: 10.1093/gbe/evac038

Ni, P., Huang, N., Nie, F., Zhang, J., Zhang, Z., Wu, B., et al. (2021). Genome-wide detection of cytosine methylations in plant from Nanopore data using deep learning. *Nat. Commun.* 12, 5976. doi: 10.1038/s41467-021-26278-9

Niederhuth, C. E., Bewick, A. J., Ji, L., Alabady, M. S., Kim, K. D., Li, Q., et al. (2016). Widespread natural variation of DNA methylation within angiosperms. *Genome Biol.* 17, 194. doi: 10.1186/s13059-016-1059-0

Ong-Abdullah, M., Ordway, J. M., Jiang, N., Ooi, S.-E., Kok, S.-Y., Sarpan, N., et al. (2015). Loss of Karma transposon methylation underlies the mantled somaclonal variant of oil palm. *Nature* 525, 533–537. doi: 10.1038/nature15365

Page, S. R. G., Monthony, A. S., and Jones, A. M. P. (2021). DKW basal salts improve micropropagation and callogenesis compared with MS basal salts in multiple commercial cultivars of *Cannabis sativa*. *Botany* 99, 269–279. doi: 10.1139/cjb-2020-0179

Peraza-Echeverria, S., Herrera-Valencia, V. A., and Kay, A.-J. (2001). Detection of DNA methylation changes in micropropagated banana plants using methylation-sensitive amplification polymorphism (MSAP). *Plant Sci.* 161, 359–367. doi: 10.1016/S0168-9452(01)00421-6

Peredo, E. L., Arroyo-García, R., and Revilla, M.Á. (2009). Epigenetic changes detected in micropropagated hop plants. *J. Plant Physiol.* 166, 1101–1111. doi: 10.1016/j.jplph.2008.12.015

Pusiak, R. J., Cox, C., and Harris, C. S. (2021). Growing pains: An overview of cannabis quality control and quality assurance in Canada. *Int. J. Drug Policy* 93, 103111. doi: 10.1016/j.drugpo.2021.103111

Quadrana, L., and Colot, V. (2016). Plant transgenerational epigenetics. *Annu. Rev. Genet.* 50, 467–491. doi: 10.1146/annurev-genet-120215-035254

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033

R Core Team. (2022). *R: A language and environment for statistical computing*. Available online at: https://www.R-project.org/.

Rival, A., Ilbert, P., Labeyrie, A., Torres, E., Doulbeau, S., Personne, A., et al. (2013). Variations in genomic DNA methylation during the long-term *in vitro* proliferation of

oil palm embryogenic suspension cultures. *Plant Cell Rep.* 32, 359–368. doi: 10.1007/s00299-012-1369-y

Roberts, R. J., and Murray, K. (1976). Restriction endonuclease. *CRC Crit. Rev. Biochem.* 4, 123–164. doi: 10.3109/10409237609105456

Russo, E. B., Jiang, H.-E., Li, X., Sutton, A., Carboni, A., del Bianco, F., et al. (2008). Phytochemical and genetic analyses of ancient cannabis from Central Asia. *J. Exp. Bot.* 59, 4171–4182. doi: 10.1093/jxb/ern260

Sato, M., Hosokawa, M., and Doi, M. (2011). Somaclonal variation is induced *de novo* via the tissue culture process: A study quantifying mutated cells in *saintpaulia*. *PloS One* 6, e23541. doi: 10.1371/journal.pone.0023541

Serres-Giardi, L., Belkhir, K., David, J., and Glémin, S. (2012). Patterns and evolution of nucleotide landscapes in seed plants. *Plant Cell* 24, 1379–1397. doi: 10.1105/tpc.111.093674

Smulders, M. J. M., Rus-Kortekaas, W., and Vosman, B. (1995). Tissue culture-induced DNA methylation polymorphisms in repetitive DNA of tomato calli and regenerated plants. *Theor. Appl. Genet.* 91, 1257–1264. doi: 10.1007/BF00220938

Smýkal, P., Valledor, L., Rodríguez, R., and Griga, M. (2007). Assessment of genetic and epigenetic stability in long-term *in vitro* shoot culture of pea (*Pisum sativum* L.). *Plant Cell Rep.* 26, 1985–1998. doi: 10.1007/s00299-007-0413-9

Springer, N. M., and Schmitz, R. J. (2017). Exploiting induced and natural epigenetic variation for crop improvement. *Nat. Rev. Genet.* 18, 563–575. doi: 10.1038/nrg.2017.45

Stelpflug, S. C., Eichten, S. R., Hermanson, P. J., Springer, N. M., and Kaeppler, S. M. (2014). Consistent and heritable alterations of DNA methylation are induced by tissue culture in maize. *Genetics* 198, 209–218. doi: 10.1534/genetics.114.165480

Suzuki, M., Liao, W., Wos, F., Johnston, A. D., DeGrazia, J., Ishii, J., et al. (2018). Whole-genome bisulfite sequencing with improved accuracy and cost. *Genome Res.* 28, 1364–1371. doi: 10.1101/gr.232587.117

Tanaka, K., and Okamoto, A. (2007). Degradation of DNA by bisulfite treatment. *Bioorg. Med. Chem. Lett.* 17, 1912–1915. doi: 10.1016/j.bmcl.2007.01.040

Torkamaneh, D., and Jones, A. M. P. (2022). Cannabis, the multibillion dollar plant that no genebank wanted. *Genome* 65, 1–5. doi: 10.1139/gen-2021-0016

Tourancheau, A., Mead, E. A., Zhang, X.-S., and Fang, G. (2021). Discovering multiple types of DNA methylation from bacteria and microbiome using nanopore sequencing. *Nat. Methods* 18, 491–498. doi: 10.1038/s41592-021-01109-3

Trucchi, E., Mazzarella, A. B., Gilfillan, G. D., Lorenzo, M. T., Schönswetter, P., and Paun, O. (2016). BsRADseq: screening DNA methylation in natural populations of non-model species. *Mol. Ecol.* 25, 1697–1713. doi: 10.1111/mec.13550

Urich, M. A., Nery, J. R., Lister, R., Schmitz, R. J., and Ecker, J. R. (2015). MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. *Nat. Protoc.* 10, 475–483. doi: 10.1038/nprot.2014.114

Us-Camas, R., Rivera-Solís, G., Duarte-Aké, F., and De-la-Peña, C. (2014). *In vitro* culture: an epigenetic challenge for plants. *Plant Cell Tissue Organ Cult. PCTOC* 118, 187–201. doi: 10.1007/s11240-014-0482-8

Van Gurp, T. P., Wagemaker, N. C. A. M., Wouters, B., Vergeer, P., Ouborg, J. N. J., and Verhoeven, K. J. F. (2016). epiGBS: reference-free reduced representation bisulfite sequencing. *Nat. Methods* 13, 322–324. doi: 10.1038/nmeth.3763

Waalwijk, C., and Flavell, R. A. (1978). MspI, an isoschizomer of HpaII which cleaves both unmethylated and methylated HpaII sites. *Nucleic Acids Res.* 5, 3231–3236. doi: 10.1093/nar/5.9.3231

Wang, X., Ke, L., Wang, S., Fu, J., Xu, J., Hao, Y., et al. (2022). Variation burst during dedifferentiation and increased CHH-type DNA methylation after 30 years of *in vitro* culture of sweet orange. *Hortic. Res.* 9, uhab036. doi: 10.1093/hr/uhab036

Wang, Q.-M., Wang, Y.-Z., Sun, L.-L., Gao, F.-Z., Sun, W., He, J., et al. (2012). Direct and indirect organogenesis of *Clivia miniata* and assessment of DNA methylation changes in various regenerated plantlets. *Plant Cell Rep.* 31, 1283–1296. doi: 10.1007/s00299-012-1248-6

Werner, O., Prudencio, Á.S., de la Cruz-Martínez, E., Nieto-Lugilde, M., Martínez-Gómez, P., and Ros, R. M. (2020). A cost reduced variant of epi-genotyping by sequencing for studying DNA methylation in non-model organisms. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00694

Wibowo, A. T., Antunez-Sanchez, J., Dawson, A., Price, J., Meehan, C., Wrightsman, T., et al. (2022). Predictable and stable epimutations induced during clonal plant propagation with embryonic transcription factor. *PloS Genet.* 18, e1010479. doi: 10.1371/journal.pgen.1010479

Wickham, H. (2011). ggplot2: ggplot2. *Wiley Interdiscip. Rev. Comput. Stat.* 3, 180–185. doi: 10.1002/wics.147

Xu, M., Li, X., and Korban, S. S. (2004). DNA-methylation alterations and exchanges during *in vitro* cellular differentiation in rose (*Rosa hybrida* L.). *Theor. Appl. Genet.* 109, 899–910. doi: 10.1007/s00122-004-1717-6

Yong, W.-S., Hsu, F.-M., and Chen, P.-Y. (2016). Profiling genome-wide DNA methylation. *Epigenet. Chromatin* 9, 26. doi: 10.1186/s13072-016-0075-3

Zhang, H., Lang, Z., and Zhu, J.-K. (2018). Dynamics and function of DNA methylation in plants. *Nat. Rev. Mol. Cell Biol.* 19, 489–506. doi: 10.1038/s41580-018-0016-z

Zhang, Z., Sun, Y., and Li, Y. (2020). Plant rejuvenation: from phenotypes to mechanisms. *Plant Cell Rep.* 39, 1249–1262. doi: 10.1007/s00299-020-02577-1

Zhang, D., Wang, Z., Wang, N., Gao, Y., Liu, Y., Wu, Y., et al. (2014). Tissue culture-induced heritable genomic variation in rice, and their phenotypic implications. *PloS One* 9, e96879. doi: 10.1371/journal.pone.0096879

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S. W.-L., Chen, H., et al. (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in *arabidopsis*. *Cell* 126, 1189–1201. doi: 10.1016/j.cell.2006.08.003