



**HAL**  
open science

## From methylome to integrative analysis of tissue specificity

Thomas Duge de Bernonville, Christian Daviaud, Cristian Chaparro, Jörg Tost, Stéphane Maury

► **To cite this version:**

Thomas Duge de Bernonville, Christian Daviaud, Cristian Chaparro, Jörg Tost, Stéphane Maury. From methylome to integrative analysis of tissue specificity. *Methods in Molecular Biology*, 2022, Catharanthus roseus Methods and Protocols, 2505, pp.223-240. 10.1007/978-1-0716-2349-7\_16 . cea-04346942

**HAL Id: cea-04346942**

**<https://cea.hal.science/cea-04346942v1>**

Submitted on 21 Oct 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## **From Methylome to Integrative Analysis of Tissue Specificity**

Thomas Dugé de Bernonville, Christian Daviaud, Cristian Chaparro, Jörg Tost, and Stéphane Maury

### **Abstract**

DNA methylation is the most studied epigenetic mark in both plants and animals. The gold standard for assaying genome-wide DNA methylation at single-base resolution is whole-genome bisulfite sequencing (WGBS). Here, we describe an improved procedure for WGBS and original bioinformatic workflows applied to unravel tissue-specific variations of the methylome in relation to gene expression and accumulation of secondary metabolites in the medicinal plant *Catharanthus roseus*.

### **Key words**

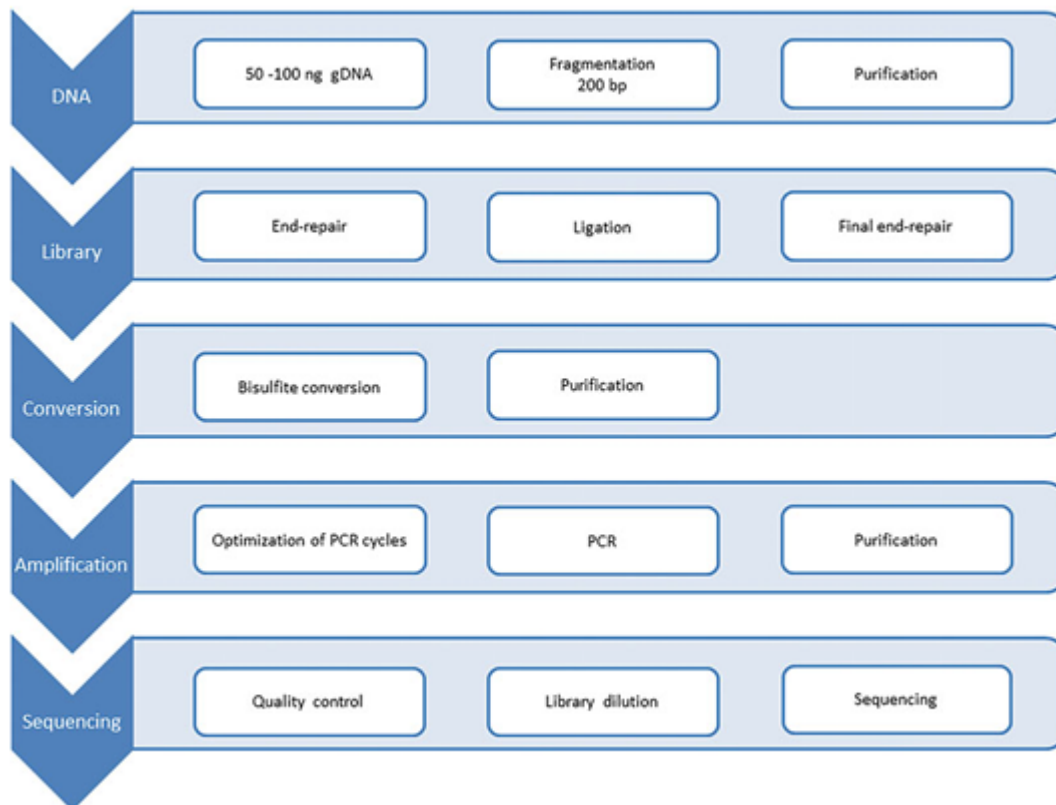
*Catharanthus roseus*, DNA methylation, Ovation Methyl-seq, Methylome pipeline, Whole-Genome Bisulfite Sequencing (WGBS)

## 1 Introduction

DNA methylation is the most studied epigenetic mark involved in gene regulation and silencing of transposable elements in plants. To assess DNA methylation profiles at cytosine level resolution, specific methods have been developed and the most widely used is whole genome bisulfite sequencing (WGBS) or MethylC-seq. WGBS consists of the fragmentation of genomic DNA, adapter ligation, bisulfite conversion, and limited amplification using adapter specific PCR primers. While initially several micrograms of DNA were required to perform WGBS, the replacement of electrophoretic steps and gel extraction by magnetic beads has enabled the creation of libraries suitable for sequencing from ~100 ng of input material [1]. Treatment of genomic DNA with sodium bisulfite modifies unmethylated cytosines into uracils that will be detected as thymines by sequencing, while methylated cytosines remain unaffected and are detected as cytosines. To assess the bisulfite conversion efficiency, DNA of the bacteriophage  $\lambda$  is spiked into the reaction. Mapping the reads against the bisulfite-converted genome of the phage and counting any remaining cytosines allows estimation of the conversion rates and identification of problems during bisulfite conversion.

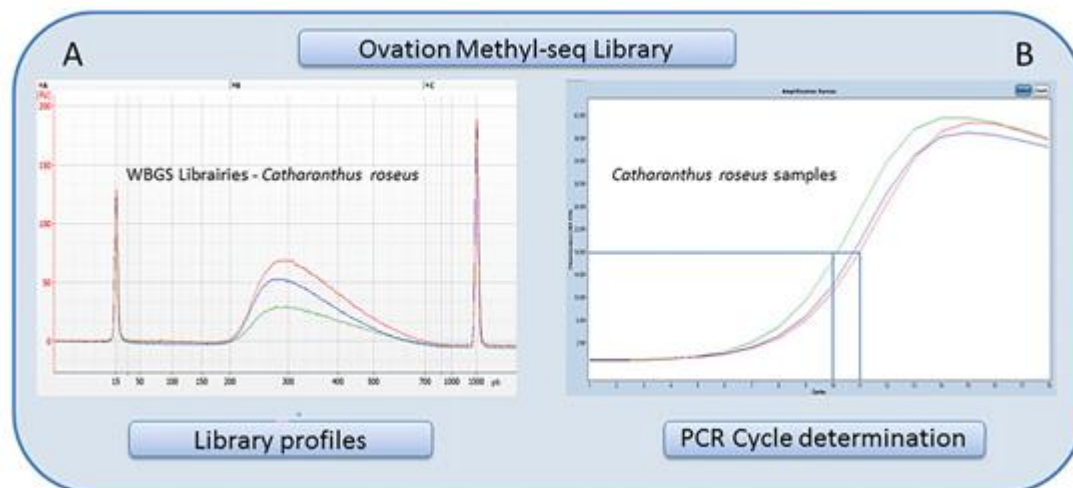
Once the WGBS libraries have been sequenced there are a couple of crucial points to be taken into account to identify methylated cytosines. The sequencing process produces errors and reads that include barcodes or sequencing adaptors that have to be cleaned. The bisulfite conversion has converted the cytosines to thymines, thus using naive alignment software that expects a matching genome will not work. This has created the need for specialized alignment software capable of taking into account the reads containing both converted and unconverted cytosines. A large number of programs have been developed to perform the quality control, pre-processing steps (such as adaptor, barcode, and quality score trimming), mapping of the reads to a bisulfite converted reference genome, scoring of DNA methylation levels (count statistics), and identification of differentially methylated cytosines (DMCs) and differentially methylated regions (DMRs) [2]. WGBS has been widely used for the methylome-wide analysis of a large number of organisms and plants, trees as well as human tissues, and ~90–95% of the cytosines present in the genome are generally covered [3–10].

The WGBS protocol using conventional bisulfite sequencing described in this protocol is based on the commercial Ovation<sup>®</sup> UltraLow Methyl-seq DR Multiplex System (Tecan Genomics/ NuGen) and has been updated and modified from previous publications [11, 12]. The workflow of the library preparation protocol is depicted in Fig. 1 and follows the classical library preparation protocol, in which methylated adaptors are ligated to the fragmented DNA prior to bisulfite conversion. One hundred ng of genomic DNA is fragmented to a size of approximately 200 base pairs (bp), purified and methylated adaptors compatible with sequencing on an Illumina MiSeq/HiSeq/NovaSeq instrument are ligated. The resulting DNA library is purified and bisulfite converted. A qPCR assay determines the optimal number of PCR amplification cycles (about 10–12 cycles) required to obtain a high-diversity library with minimal duplicate reads prior to the actual library amplification. Due to the specific proprietary structure and sequence of the adaptors employed in the protocol, a dedicated sequencing primer has to be added during cluster generation on the cBOT of Illumina HiSeq sequencers, which will be used as the sequencing primer for read 1. The index read as well as the second read in pair-end mode will be using standard Illumina primers. Following sequencing using 2 X 150 bp reads, the data is quality controlled and mapped to the genome to obtain the percentage of DNA methylation at all cytosines.



**FIG 1.** Workflow of the whole-genome bisulfite sequencing protocol using the Ovation Low Input Methyl-seq Kit. (Reproduced from Ref. [12]. See the text for a detailed description)

The described protocol has been applied to the analysis of the methylome of the medicinal plant *Catharanthus roseus* (Fig. 2) followed by original bioinformatic workflows (Fig. 3) [13]. This plant is able to produce many monoterpene indole alkaloids, with some of them displaying pharmacological relevant activities. Such compounds are produced by many consecutive enzymatic steps, encoded by genes whose expression is known to be regulated by transcription factors [14]. Some of these alkaloids display a spatiotemporal-specific accumulation pattern and how this specific distribution is achieved is not well understood. Differential methylation in promoters and gene bodies could be involved in this mechanism and was therefore investigated in different *C. roseus* samples reflecting different tissues and experimental conditions (Fig. 4a) [13]. Because of the incompleteness of *C. roseus* genome in 2015, WGBS reads were also mapped to gene models  $\pm$  5Kb (Fig. 4b). Using original bioinformatic workflows (Fig. 3), DMCs and DMRs have been called from pairwise sample comparison (Fig. 4c). Gene methylation profiles were compared to gene expression profiles using clustering methods (Fig. 4d) and functional annotations allowed the identification of processes potentially regulated by DNA methylation in *C. roseus*.

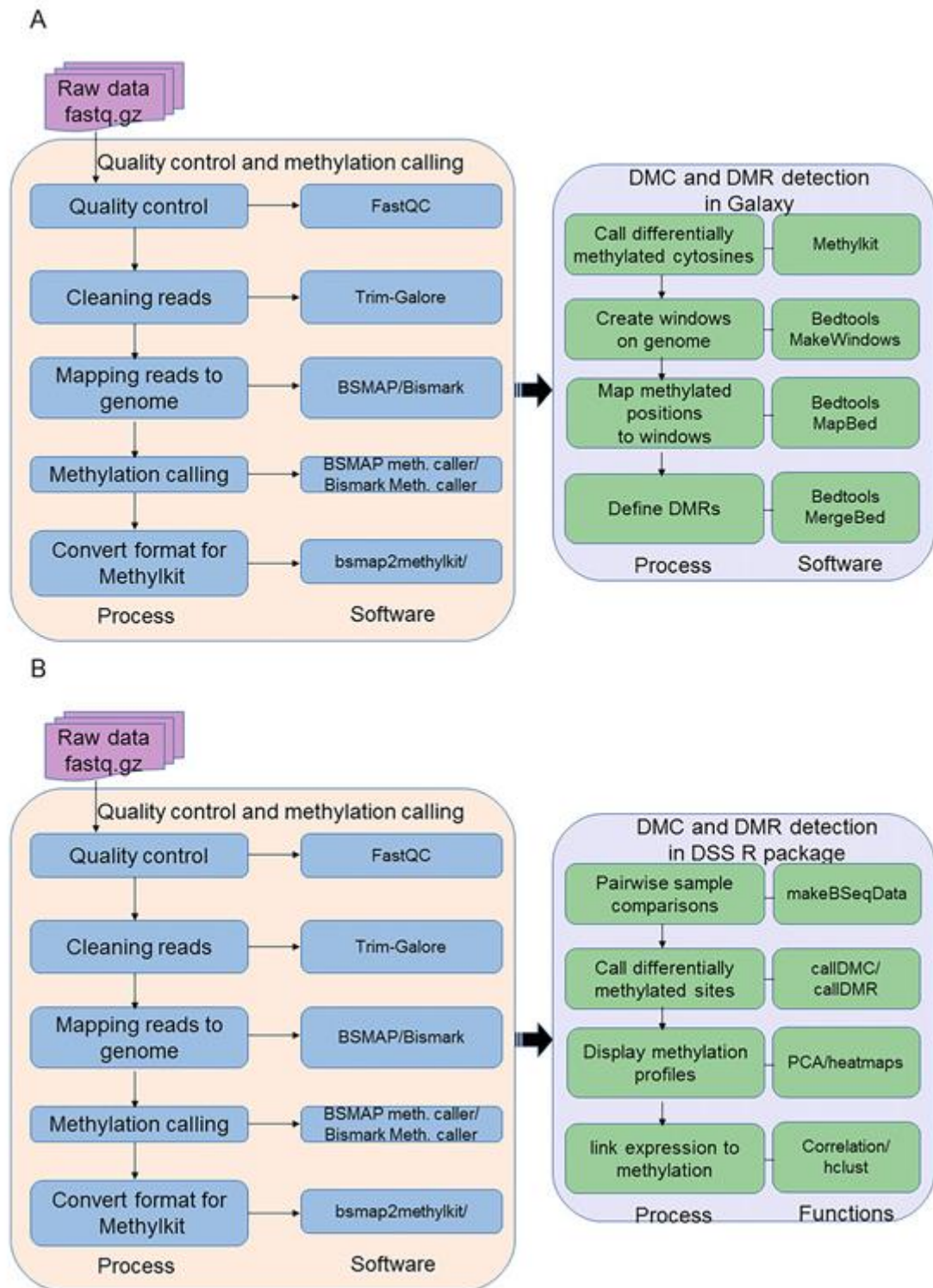


**FIG. 2** WGBS library construction for samples from *C. roseus*. The green, red and blue curves represent three libraries made from 100 ng DNA extracted from tissues of *Catharanthus roseus*. (a) A representative Bioanalyzer profile using DNA 1000 Chips. The x axis shows the size distribution of the amplification products in base pairs, and the y axis shows units of fluorescence. The average library size is approximately 290 base pairs. (b) Determination of the optimal number of PCR cycles. Relative fluorescence is plotted against the number of PCR cycles. For the three libraries (green, red, and blue curves) the optimal number of cycles is respectively, 10, 11, and 11 cycles. (Modified from Ref. [12])

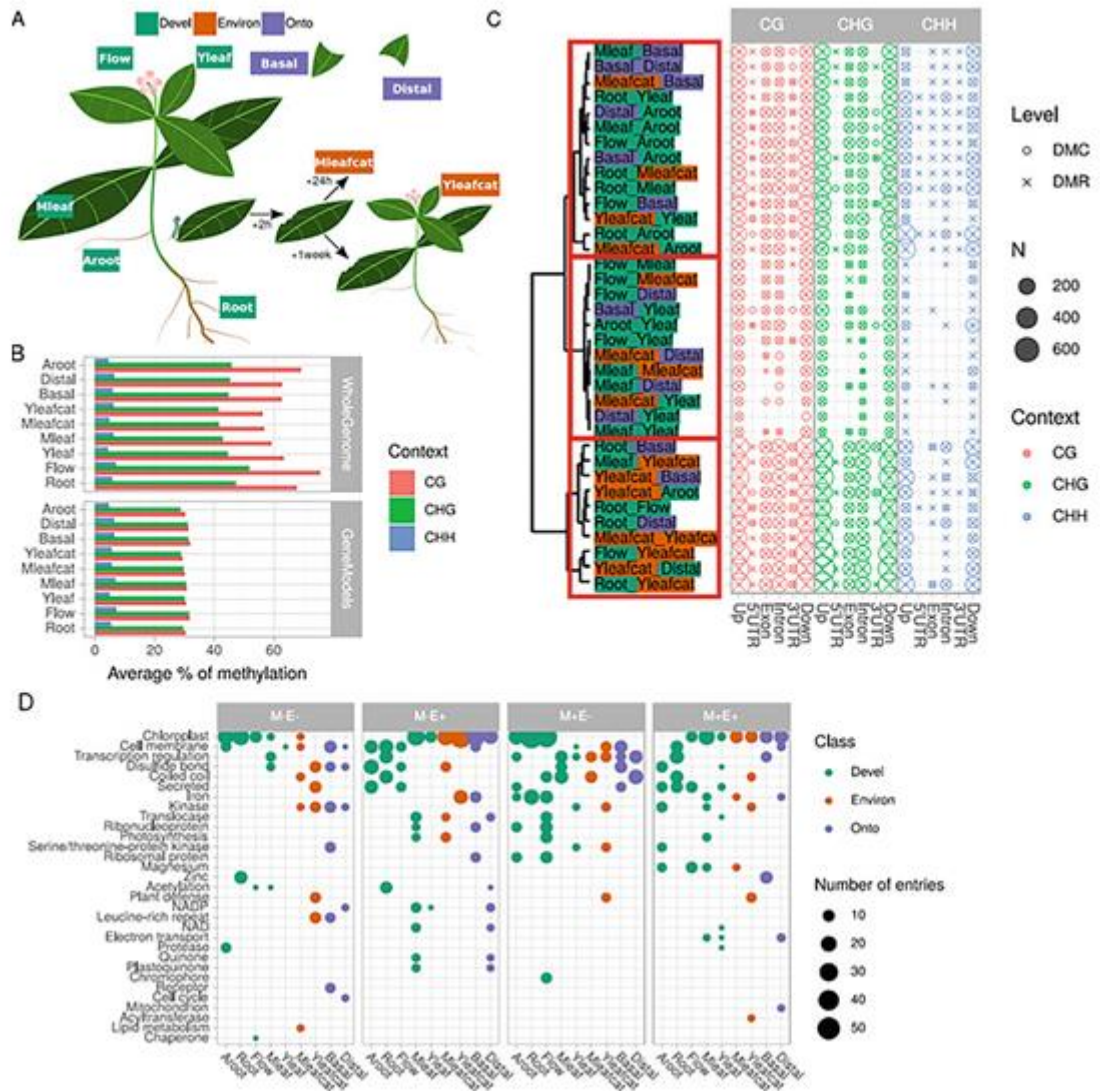
## 2 Materials

### 2.1 Reagents and Consumables

1. 100 ng of genomic DNA (see Note 1).
2. Ultralow Methy-Seq System, 96 rxn (NuGEN) containing: End Repair Buffer Mix (ER1 ver 8), End Repair Enzyme Mix (ER2ver4), EndRepairEnhancer(ER3ver2), LigationBuffer Mix (L1 ver 4), Ligation Adaptor Plate (L2V11DR-BC1 to L2V11DR-BC96), Ligation Enzyme Mix (L3 ver 4), Final Repair Buffer Mix (FR1 ver 4), Final Repair Enzyme (FR2), Amplification Primer Mix (P2 ver 8), Amplification Enzyme Mix (P3 ver 3), 25  $\mu$ M Sequencing Primer (MetSeq Primer 1), nuclease-free water (D1), and Agencourt RNACleanXPBeads.
3. Unmethylated cl857 Sam7 Lambda DNA.
4. Bisulfite Conversion Kits such as EpiTect FAST DNA Bisulfite kit (Qiagen) containing Bisulfite Solution, DNA Protect Buffer, RNase-Free Water, MinElute DNA Spin Columns, collection Tubes (2 mL), BL Buffer, BW Buffer, BD Buffer, EB Buffer, and Carrier RNA.
5. 100% and 70% Ethanol.
6. 1 x TE Buffer: 10 mM Tris-HCL, 1 mM EDTA, pH 8.0.
7. EvaGreen dye 20x .
8. Covaris tube: 6 x 16 mm Round bottom glass, AFA fiber and cap system (100  $\mu$ L).
9. 0.2 mL and 1.5 mL tubes "PCR clean" with lid.
10. 384 wells plate and seal for qPCR.
11. Library quantification kit (e.g., KAPA Library Quantification Kit, Roche).



**FIG. 3** Two original workflows for bioinformatics analysis of methylome data. (a) Methykit workflow in Galaxy, (b) DSS workflow in R.



**FIG. 4** Tissue-specific analysis of methylome data in *C. roseus*. (a) Nine samples reflecting developmental variation within distinct organs (leaf, flower, root, or adventitious root) or during leaf ontogenesis (young versus mature; or basal versus distal) and environmental effects on leaf after direct (mature leaf 24 h post folivory) or indirect (distance young leaf 1 week post folivory) biotic attacks were chosen to characterize whole-genome cytosine methylation profiles. Flow Flower, Aroot Adventitious root, Mleaf Mature leaf, Yleaf Young leaf, Basal Basal part of leaf, Distal Distal part of leaf, Mleafcat Mature leaf post folivory by caterpillars, Yleafcat Young leaf from subjected plants 1 week after folivory on mature leaves. (b) Average percentage of cytosine methylation in the whole-genome or gene model targeted approaches. (c) Number of DMC and DMR in all 36 pairwise comparisons, for each context and gene model feature. Comparisons were clustered using a Euclidean distance calculated on the number of DMCs/DMRs and the Ward agglomeration criterion. (d) Most represented Uniprot keywords for differentially expressed and methylated (DEM) genes for each sample and methylation/expression profile. For each sample, all nine comparisons were made to identify gene sets following one of the four profiles: M-E-, hypomethylated and decreased expression; M-E+, hypomethylated and overexpressed; M+E-, hypermethylated and decreased expression; M+E+, hypermethylated and over expressed. Each corresponding Uniprot entry was mapped to Uniprot keywords, which were then summarized for each sample. (Modified from Ref. [13])

## **2.2 Equipment**

1. Magnetic rack for 1.5 mL tubes (e.g., DiaMag02, Diagenode).
2. Thermocycler.
3. qPCR thermocycler.
4. Ultrasonicator (e.g., E220, Covaris).
5. Bioanalyser (e.g., Bioanalyser 2100 instrument, Agilent including a DNA 1000 kit).
6. Centrifuge.
7. Laminar Flow Cabinets (optional, see Note 2).

## **2.3 Genome Resource**

1. High-quality genome assembly (e.g., from NCBI, EBI or Phytozome).
2. Comprehensive annotation in gff format, including the following features: gene, upstream, downstream UTRs, exons, and introns. If possible, include TSS.
3. Expression data (or other relevant data sets such as metabolomic data) for the same or similar samples.

## **2.4 Galaxy or R Packages**

1. Galaxy server, which can be a public instance or a locally installed one. Galaxy is an open-source, web-based platform for data-intensive biomedical research [15]. R packages are collections of functions and data sets developed by the community.
2. FastQC and Trim Galore packages for quality control and cleaning of reads.
3. BSMAP[16] and/or Bismark [17] for mapping of sequencing reads from WGBS.
4. Methylkit [18] or DSS R package [19] for differential methylation analysis.
5. Data table for fast operations on large tables.
6. EdgeR for differential gene expression.
7. R Library (ggplot2).
8. R Library (ggdendro).
9. R Library (FactoMineR).

## **3 Methods**

### **3.1 Ovation Methyl seq Workflow**

#### **3.1.1 DNA Fragmentation**

1. Dilute 100 ng of genomic DNA into 50  $\mu$ L of 1 x TE buffer.
2. Add 0.25% w/w of unmethylated Lambda DNA (e.g., 0.25 ng of Lambda DNA for 100 ng DNA).
3. Transfer the sample to a Covaris tube.
4. Fragment the DNA using a Covaris ultrasonicator. For a 200 bp target length, use the following Covaris parameters: 36 W of power, 10% duty cycle, 175 of peak incident power, 200 cycles per burst. Fragmentation is conducted two times 60 s (spin between the two runs of fragmentation) at 8° C using 50  $\mu$ L of DNA.
5. After fragmentation, briefly spin the Covaris tube and transfer the entire sample to a new 0.2 mL tube.



### 3.1.2 DNA Purification

1. Thirty minutes before proceeding take the Agencourt RNA Clean XP beads and nuclease-free water out of the fridge and leave to warm up at room temperature.
2. Resuspend the RNAClean XP beads by vortexing the tube. Ensure beads are fully resuspended before adding to the sample.
3. Prepare afresh 70% ethanol wash solution (1.5 mL per sample).
4. Add 90  $\mu$ L (1.8 volumes) of the bead suspension to the 50  $\mu$ L of fragmented DNA and mix by pipetting 10 times.
5. Incubate at room temperature for 10 min.
6. Place the tube into the magnetic rack and incubate for 5 min at room temperature to completely clear the solution.
7. Carefully remove 130  $\mu$ L of the binding buffer and discard it (see Note 3).
8. Add 200  $\mu$ L of freshly prepared 70% ethanol, wash and leave to stand for 30 s.
9. Remove and discard the 70% ethanol using a pipette.
10. Repeat steps 8 and 9 one more time, for a total of two 70% ethanol washes (see Note 4).
11. Air-dry the bead pellets on the magnetic rack for 5–10 min at roomtemperature. Inspect eachtube carefully to ensure that all the ethanol has evaporated.
12. Add 15  $\mu$ L room temperature nuclease-free water (provided in the kit) to the dried beads. Mix thoroughly to ensure all the beads are resuspended.
13. Place the tube into the magnetic separation rack and incubate for 5 min at room temperature to completely clear the solution.
14. Carefully transfer 13  $\mu$ L of the eluate into a new 0.2 mL tube ensuring to carry over as few beads as possible and place on ice.

### 3.1.3 End Repair

1. Thaw the End Repair Buffer (ER1) of the Ovation Ultralow Methyl-Seq DR Multiplex System.
2. Prepare the End-Repair master mix by combining in a 0.2 mL tube 2  $\mu$ L of End Repair Buffer (ER1), 0.5  $\mu$ L of End Repair Enzyme mix (ER2 ver 4), and 0.5  $\mu$ L of End Repair Enhancer (ER3 ver 2).
3. Add 3  $\mu$ L of the End Repair master mix to the fragmented DNAsample (from Subheading 3.1.2). Mix by pipetting, cap, and briefly spin down the tubes and place on ice.
4. Incubate the sample in a thermal cycler at 25° C for 30 min followed by incubation at 70°C for 10 min and cool down to 4° C.
5. Remove the tubes from the thermal cycler, briefly centrifuge, and place on ice.

### 3.1.4 Ligation

1. Choose and thaw one adaptor per sample from the Ligation Adaptor Mixes (L2V11DR-BC1 through L2V11DR-BC96). Thaw the Ligation Buffer Mix (L1 ver4) on ice. Thaw the nuclease-free water (D1) at room temperature.
2. Add 3  $\mu$ L of L2 Ligation Adaptor Mix to each sample (from step 5 of Subheading 3.1.3). Mix thoroughly by pipetting.
3. Prepare a Ligation master mix by combining in a 0.2 mL tube 4.5  $\mu$ L of nuclease-free water (D1), 6  $\mu$ L of Ligation Buffer Mix, and 1.5  $\mu$ L of Ligation Enzyme Mix. Mix by pipetting slowly, without introducing bubbles, briefly spin down and place on ice (see Note 5).
4. Add 12  $\mu$ L of Ligation Master Mix to each reaction tube. Mix thoroughly by pipetting slowly, spin, and place on ice. Proceed immediately to the next step.

5. Incubate the sample in a thermal cycler at 25° C for 30 min followed by incubation at 70° C for 10 min and cool down to 4° C.
6. Remove the tubes from the thermal cycler, briefly spin down, and place on ice.
7. Purify the ligation products as described in steps 1–11 of Subheading 3.1.2 with the exception of adding 45 µL (1.5 volumes) of the bead suspension to the 30 µL of sample and mix by pipetting 10 times.
8. Add 16 µL nuclease-free water (green: D1) equilibrated to room temperature to the dried beads. Mix thoroughly to ensure all the beads are resuspended.
9. Transfer the tube into the magnetic separation racks and incubate for 5 min at room temperature to completely clear the solution.
10. Carefully transfer 15 µL of the eluate into a new PCR tube ensuring to carry over as few beads as possible and place on ice.

### *3.1.5 Final Repair*

1. Thaw the Final Repair Buffer Mix (Purple: FR1 VER 4) on ice.
2. Make a Final Repair master mix by combining 4.5 µL of Final Repair Buffer Mix (FR1) and 0.5 µL of Final Repair Enzyme Mix (FR2) in a 0.2 mL capped tube. Mix by pipetting, spin down briefly, and place the tube on ice.
3. Add 5 µL of Final Repair master mix to each sample (from Subheading 3.1.4).
4. Mix thoroughly by pipetting slowly, spin, and place on ice.
5. Incubate the sample in a thermal cycler at 60° C for 10 min and cool down to 4° C.
6. Remove the tubes from the thermal cycler, spin down briefly, and place on ice.

### *3.1.6 Bisulfite Conversion (see Note 6)*

1. Thaw the DNA to be used in the bisulfite reactions. Make sure the Bisulfite Solution is completely dissolved (see Note 7).
2. Add 85 µL of Bisulfite Solution and 35 µL of DNA Protect Buffer to each sample (20 µL from Subheading 3.1.5). Mix the bisulfite reaction thoroughly and gently spin down. Place the tube at room temperature (15–25° C) (see Note 8).
3. Place the reaction in a thermal cycler and incubate the sample use the following program: 5 min at 95° C, 20 min at 60° C, 5 min at 95° C, 40 min at 60° C, 5 min at 95° C, 45 min at 60° C, and hold at 20° C.

### *3.1.7 Cleanup of Bisulfite Converted DNA*

1. Briefly centrifuge the tubes containing the bisulfite reactions, and then transfer the totality of the bisulfite reactions into new 1.5 mL microcentrifuge tubes.
2. Add 310 µL of freshly prepared BL Buffer containing 10 µg/ mL carrier RNA to each sample. Mix the solutions by vortexing and then centrifuge briefly.
3. Add 250 µL ethanol (96–100%) to each sample. Mix the solutions by pulse vortexing for 15 s, and centrifuge briefly.
4. Place a MinElute DNA spin column and a collection tube into a suitable rack and transfer the entire mixture from each tube into the corresponding MinElute DNA spin column.
5. Centrifuge the spin column at 15,000 x g for 1 min at room temperature. Discard the flow-through, and place the spin column back into the collection tube.

6. Add 500  $\mu$ L of BW Buffer (wash buffer) to each spin column, and centrifuge at maximum speed for 1 min at room temperature. Discard the flow-through, and place the spin column back into the collection tube.

7. Add 500  $\mu$ L of BD Buffer (desulfonation buffer) to each spin column (see Notes 9 and 10), and incubate for 15 min at room temperature.

8. Centrifuge the spin columns at 15,000  $\times$  g for 1 min at room temperature. Discard the flow-through, and place the spin column back into the collection tube.

9. Add 500  $\mu$ L of BW Buffer to each spin column and centrifuge at 15,000  $\times$  g for 1 min at room temperature. Discard the flow through and place the spin column back into the collection tube.

10. Add 500  $\mu$ L of BW Buffer to each spin column and centrifuge at 15,000  $\times$  g for 1 min at room temperature. Discard the flow through and place the spin column back into the collection tube.

11. Add 250  $\mu$ L of ethanol (96–100%) to each spin column and centrifuge at 15,000  $\times$  g for 1 min at room temperature.

12. Place the spin column into a new 2 mL collection tube and centrifuge at 15,000  $\times$  g for 1 min to remove any residual liquid.

13. Place the spin column into a new 1.5 mL tube. Add 22  $\mu$ L of the EB Buffer (elution buffer) directly onto the center of each spin-column membrane and close the lid gently.

14. Incubate the spin column at room temperature for 1 min.

15. Centrifuge for 1 min at 15,000  $\times$  g at room temperature to elute the DNA.

16. Transfer 20  $\mu$ L of the eluate into a new 0.2 mL tube and place on ice (see Note 11).

### *3.1.8 Determination of Optimal Amplification Cycles by qPCR*

1. Thaw the Amplification Primer Mix (P2 ver 5) and Amplification Enzyme Mix (P3) on ice.

2. Transfer 4  $\mu$ L of converted DNA (from step 16 of Subheading 3.1.7) into a 384 wells plate.

3. Prepare a qPCR master mix by combining 1  $\mu$ L of Amplification Primer Mix (P2 ver 5), 0.5  $\mu$ L of EvaGreen 20 X and 4.5  $\mu$ L of Amplification Enzyme Mix (P3) in a 0.5 mL capped tube. Mix by pipetting, spin down briefly, and place on ice.

4. Add 6  $\mu$ L of qPCR master mix to 4  $\mu$ L of the library and seal the plate.

5. Centrifuge at 4000  $g$  for 1 min at room temperature and place the tube on ice.

6. Place the reaction into a qPCR thermocycler using the following program: 2 min at 95° C, 35 cycles of 15 s at 95° C, 1 min at 60° C, 30 s at 72° C, and optical reading of the plate in SYBR green mode. Finally, hold at 40° C.

7. Determine the optimal number of cycles needed for sufficient amplification without over-amplifying the library (N): plot linear Rn (see Note 12) versus cycle and determine the cycle number that corresponds to 1/2 of the maximum fluorescent intensity (Fig. 2).

### *3.1.9 PCR Amplification*

1. Thaw the Amplification Primer Mix (P2 ver 5) and Amplification Enzyme Mix (P3) on ice.

2. Make a PCR master mix by combining 4  $\mu$ L of Amplification Primer Mix (P2 ver 5) and 20  $\mu$ L of Amplification Enzyme Mix (P3) in a 1.5 mL capped tube. Mix by pipetting, spin down the tube, and place on ice.

3. Add 24  $\mu$ L of PCR master mix to each sample (16  $\mu$ L from Subheading 3.1.7).

4. Place the reaction in a pre-heated thermal cycler using the following program: 2 min at 95° C, N cycles (N = as determined in Subheading 3.1.8) of 15 s at 95° C, 1 min at 60° C, and 30 s at 72° C. Finally, hold at 4° C.

5. Remove the tube from the thermal cycler, spin down briefly, and place on ice.

6. Purify the amplification products as described in steps 1–11 of Subheading 3.1.2 with the exception of adding 50  $\mu$ L (1 volume) of the bead suspension to the 50  $\mu$ L of sample and mix by pipetting 10 times.

7. Add 20  $\mu$ L of the nuclease-free water (green: D1) equilibrated to room temperature to the dried beads. Mix thoroughly to ensure all the beads are resuspended.

8. Transfer the tube into the magnetic separation racks and incubate for 5 min at room temperature to completely clear the solution.

9. Carefully transfer 18  $\mu$ L of the eluate into a new PCR tube ensuring to carry over as few beads as possible and place on ice.

### *3.1.10 Quantitative and Qualitative Assessment of the Library*

1. Thirty minutes before proceeding, take the DNA1000 kit reagents out of the fridge and leave to equilibrate to room temperature.

2. Run 1  $\mu$ L of the libraries on a Bioanalyzer DNA1000 Chip.

3. Profiles should display a peak at approximately 300 bp (270 and 420 bp) as shown in Fig. 2 corresponding to 150–200 bp inserts (see Note 13).

4. Perform a qPCR quantification of the library preparation using the KAPA Library Quantification Kit. Prepare a solution of Elution Buffer/0.001% Tween20 and two dilutions of each sample at 1/10000 and 1/100000.

5. In a qPCR plate, add 6  $\mu$ L of KAPA SyBR qPCR Master Mix and 4  $\mu$ L of the library preparation or standard (included in the KAPA Library Quantification Kit). Each sample should be analyzed in triplicate. Seal the plate and centrifuge it at 4000 *g* for 1 min at room temperature.

6. Place the plate into a qPCR thermo cycler and perform the following program: 5 min at 95° C, 40 cycles of 30 s at 95° C, 45 s at 60° C, and optical reading of the plate in SYBR green mode. Finally, hold at 40° C.

7. Determine the Ct (Cycle Threshold) using the Abs Quant/2nd Derivative Max for each sample. The Ct should be between 8 and 26 cycles. Determine the concentration of each sample using the Ct and the standard curve. The concentration of each sample should be in the range of the standard curve, i.e., between 0.0002 pM and 20 pM. Calculate the library concentration using this formula: Library concentration in nM = [Conc qPCR] X dilution factor X 452 (corresponding to the length of the standard included in the KAPA Library Quantification Kit)/average size of the library preparation (see Note 14).

### **3.2 Sequencing**

1. Sequencing is performed using standard Illumina sequencing protocols in 2 X 150 bp following the instructions provided by Illumina.

2. Prepare a multiplex equimolar pool of samples standardized with 5% of PhiX at 2.6 nM in EB/0.001% Tween 20 into 30  $\mu$ L final volume (see Note 15).

3. The design of the Ovation Ultralow Methyl-Seq DR Multiplex System requires the use of a custom read 1 sequencing primer, called MetSeq Primer 1 that is included in this kit at a concentration of 100  $\mu$ M. 2.4  $\mu$ L of this primer is added manually during the Cluster Generation Process on a cBOT for sequencing on a HiSeq 4000. The standard primers provided in the Illumina sequencing kit are appropriate for read 2 and for sequencing the DR barcodes.

### **3.3 Processing of Sequencing Data (Fig. 3)**

### 3.3.1 Quality Control of Sequencing Reads and Cleaning

The quality control of the sequencing files in FASTQ can be done, before and after cleaning the files.

1. Use FastQC (galaxy version 0.72) on each sequence file to obtain an overview of the quality of the sequences.

2. MultiQC (galaxy version 1.9) is used to produce a summary of all files in the experiment.

3. Cleaning of the raw data is carried out by considering reads that have over 95% of their nucleotides presenting a quality score of over 26 by using the filter by a quality tool (Galaxy version 1.0.0). Adapters, reads that are too short, and pairs that are missing one member are eliminated using TrimGalore V06.3 (<https://github.com/FelixKrueger/TrimGalore>). In case of pair-end sequences, unpaired reads can also be eliminated by using the FASTQ interlace (Galaxy version 1.1) and FASTQ deinterlace programs (galaxy version 1.1)

### 3.3.2 Alignment and Methylation Calling

1. Reads are aligned to the reference genome using Bismark (version 0.16.3) [17] with Bowtie2 (version 2.1.0) providing also global methylation percentages per context (CpG, CHG, and CHH, with H=A, T or C). Alternatively, mapping can be done with the BSMAP mapper tool in Galaxy (v1.0.0) which uses BSMAP 2.7 [16] (see Note 16).

2. Methylation is called using Bismark Meth. Extractor tool (galaxy version 0.20.0) or BSMAP methylation caller tool (v1.0.0), respectively and generate .bed files. Subsequently, the methylation call file is converted to a format suitable for Methylkit using methylextract2methylkit (V0.1.0) or bsmmap2Methylkit (V0.1.0), respectively. These scripts are run for each methylation context CpG, CHG, and CHH.

### 3.3.3 Identification of Differential Methylation: DMC and DMR (Fig. 3)

#### *Workflow 1 (Galaxy, see Note 17)*

1. Methylkit (galaxy version 0.99.2) [18] is then used to call the differentially methylated sites (DMC) by selecting those that have a minimum coverage of 10 reads and have a  $q$ value  $\leq 0.01$  (`-qvalue.cutoff`) and methylation difference  $> 25\%$  (`-meth. cutoff`) for CpG and CHG and  $> 10\%$  for CHH context.

2. To identify the DMRs, the DMCs are mapped to genomic windows and the average methylation is calculated for each of the windows. Windows are produced with Bedtools MakeWindowsBed (galaxy version 2.29.2).

#### *Workflow 2 (DSS in R, see Note 18)*

This part will be entirely performed in R. DMC and DMR will be identified with the DSS R package [19] which is able to process datasets without biological replicates. DMR are automatically detected by DSS from DMC. Regions with many statistically significant DMC are identified as DMR. Because initially intended to be used on animal methylomes, the pipeline has to be run for each C context.

1. Load the generated tabular files with BSMAP methylation caller tool: a loop using the “`fread`” function to store all tabular files into the R environment. A minimum CT coverage can be set, e.g., to retain sites covered by at least 3 reads.

2. Convert to the appropriate data structure with “makeBSseq Data”. Only columns chr, pos, X (total number of reads), and N (number of reads with methylation) will be used.
3. Statistical modeling: for each pairwise comparison, run a “DMLtest” function using the ‘smoothing’ argument to allow the use of neighboring cytosines as pseudo-replicates.
4. Detect DMCs with the “callDML” function.
5. Detect DMRs with the “callDMR” function.

### **3.4 Data Exploration (see Note 18)**

1. Combine comparisons into a single table.
2. Annotate cytosine sites.
3. Inspect methylation profiles per gene feature and context.
4. Cluster samples according to profiles.
5. Visualize the number of DMCs and DMRs.
6. Principal component analysis.
7. Correlation between gene expression and methylation level.
8. Functional exploration of DMC and DMR.

### **4 Notes**

1. Between 50 and 200 ng of input, DNA allows obtaining high-diversity libraries with comprehensive genomic coverage. We have successfully sequenced samples with input material as limited as 30 ng. However, below this amount, the duplicate rates increase significantly and for sufficient coverage, extensive sequencing is required. We therefore recommend a minimum input of 50 ng and ideally 100 ng for routine use.
2. For all procedures prior to PCR amplification, it is recommended to work in Laminar Flow Cabinets and use filter tips. After PCR, it is recommended to work in a physically separated post-PCR room.
3. The beads should not disperse and should stick to the walls of the tubes to avoid loss during the wash process. Significant loss of beads at this stage will impact the amount of purified DNA, so ensure beads are not removed with the binding buffer or during the wash steps.
4. With the final wash, it is critical to remove as much of the ethanol as possible. If necessary, use extra pipetting steps and allow excess ethanol to collect at the bottom of the tubes after removing most of the ethanol in the first pipetting step.
5. The L1 Ligation Buffer Mix is very viscous. Pipet this reagent slowly.
6. We have evaluated several commercial bisulfite conversion kits and all kits tested have been found to be compatible with the library preparation protocol. However, the kit cited within this protocol allowed us to obtain the most consistent results and the best yields after bisulfite conversion with excellent conversion rates.
7. If necessary, heat the Bisulfite Solution to 60° C and vortex until all precipitates are dissolved. Do not place dissolved bisulfite solution on ice.
8. DNA Protect Buffer should turn from green to blue after addition to the DNA–Bisulfite Solution mixture, indicating sufficient mixing and correct pH for the bisulfite conversion reaction.
9. If there are precipitates in the BD Buffer, avoid transferring them to the spin column.
10. To avoid acidification from carbon dioxide in the air, it is important to close the bottle containing the BD Buffer immediately after use and close the lids of the spin columns before incubation.
11. Proceed immediately or store purified DNA at - 20° C.

12.  $R_n$  is the fluorescence of the reporter dye divided by the fluorescence of a passive reference dye.

13. Slightly bigger inserts have not been found to negatively impact the sequencing of the library. For significantly larger inserts, this might however have an influence on library diversity and as inserts are penalized during cluster generation on the cBot.

14. 452 bp is the average size of the standard included in the KAPA library quantification kit. The average size of the library preparation is obtained in the Bioanalyzer analysis.

15. Depending on the software used on the sequencing machine, higher amounts of PhiX might be required. Prior to the HiSeq Control software (HCS) V3.4.0., a minimum of 20% PhiX was required.

16. Mapping using BSMAP/Bismark can be done on a reduced representation of the genome consisting of the annotated genes including a flanking region. BSMAP needs the shortest run time and yields the highest precision while Bismark requires the smallest amount of memory and yields precision and high numbers of uniquely mapped reads [20].

17. This galaxy workflow is available at IHPE Galaxy Perpignan, ([https://genome.univ-perp.fr/galaxy/Galaxy-Workflow\\_DMR\\_from\\_Methylkit\\_differential\\_methylation.ga](https://genome.univ-perp.fr/galaxy/Galaxy-Workflow_DMR_from_Methylkit_differential_methylation.ga)).

18. Scripts and data to reproduce results presented here are available as a Github ([https://github.com/EA2106-Universite\\_Francois-Rabelais/epigenomics](https://github.com/EA2106-Universite_Francois-Rabelais/epigenomics)). Examples of analysis are shown in Fig. 4, adapted from [13].

### **Acknowledgments**

The protocol for WGBS analysis has been set-up in the laboratory of J. Tost in the framework of the ANR-BMBF funded project “Epigenomics of Parkinson’s Disease” (EpiPD, ANR-13-EPIG 0003-05). The research on *C. roseus* was funded by the “Région Centre Val de Loire” (France, ABISAL, and CatharSIS grants) for the University of Tours (V. Courdavault). Part of the methylome analysis was also developed in the frame of the ANR funded project EPITREE (ANR-17-CE32-0009-01; [https://www6.inrae.fr/epitree-project\\_eng/](https://www6.inrae.fr/epitree-project_eng/)) at the University of Orléans (S. Maury) and the CEA (J. Tost) in collaboration with the Plateforme Epigénomique Environnementale, IHPE Perpignan (<https://bioinfo.univ-perp.fr/>) providing computational infrastructure and bioinformatics support (C. Chaparro).

## References

1. Urich MA, Nery JR, Lister R, Schmitz RJ, Ecker JR (2015) MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. *Nat Protoc* 10:475–483
2. Adusumalli S, Mohd Omar MF, Soong R, Benoukraf T (2015) Methodological aspects of whole-genome bisulfite sequencing analysis. *Brief Bioinform* 16:369–379
3. Lister R, Pelizzola M, Downen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsall L, Antosiewicz Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462:315–322
4. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR (2008) Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* 133:523–536
5. Lister R, Pelizzola M, Kida YS, Hawkins RD, Nery JR, Hon G, Antosiewicz-Bourget J, O'Malley R, Castanon R, Klugman S, Downes M, Yu R, Stewart R, Ren B, Thomson JA, Evans RM, Ecker JR (2011) Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* 471: 68–73
6. Li Y, Zhu J, Tian G, Li N, Li Q, Ye M, Zheng H, Yu J, Wu H, Sun J, Zhang H, Chen Q, Luo R, Chen M, He Y, Jin X, Zhang Q, Yu C, Zhou G, Huang Y, Cao H, Zhou X, Guo S, Hu X, Li X, Kristiansen K, Bolund L, Xu J, Wang W, Yang H, Wang J, Li R, Beck S, Zhang X (2010) The DNA methylome of human peripheral blood mono nuclear cells. *PLoS Biol* 8:e1000533
7. Chalhoub B, Denoeud F, Liu S, Parkin IA, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, Correa M, Da Silva C, Just J, Falentin C, Koh CS, Le Clainche I, Bernard M, Bento P, Noel B, Labadie K, Alberti A, Charles M, Arnaud D, Guo H, Daviaud C, Alamery S, Jabbari K, Zhao M, Edger PP, Chelaifa H, Tack D, Lassalle G, Mestiri I, Schnell N, Le Paslier MC, Fan G, Renault V, Bayer PE, Golicz AA, Manoli S, Lee TH, Thi VH, Chalabi S, Hu Q, Fan C, Tollenaere R, Lu Y, Battail C, Shen J, Side bottom CH, Wang X, Canaguier A, Chauveau A, Berard A, Deniot G, Guan M, Liu Z, Sun F, Lim YP, Lyons E, Town CD, Bancroft I, Wang X, Meng J, Ma J, Pires JC, King GJ, Brunel D, Delourme R, Renard M, Aury JM, Adams KL, Batley J, Snowdon RJ, Tost J, Edwards D, Zhou Y, Hua W, Sharpe AG, Paterson AH, Guan C, Wincker P (2014) Plant genetics. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* 345:950–953
8. Lyko F, Foret S, Kucharski R, Wolf S, Falckenhayn C, Maleszka R (2010) The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS Biol* 8:e1000506
9. Guo JU, Su Y, Shin JH, Shin J, Li H, Xie B, Zhong C, Hu S, Le T, Fan G, Zhu H, Chang Q, Gao Y, Ming GL, Song H (2014) Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nat Neurosci* 17:215–222
10. Sow MD, Allona I, Ambroise C, Conde D, Fichot R, Gribkova S, Jorge V, Le-Provost G, Pâques L, Plomion C, Salse J, Sanchez Rodriguez L, Segura V, Tost J, Maury S (2018) Epigenetics in forest trees. In: *Advances in botanical research*, vol 88. Elsevier, Amsterdam, pp 387–453
11. Daviaud C, Renault V, Mauger F, Deleuze JF, Tost J (2018) Whole-genome bisulfite sequencing using the ovation (R) ultralow methyl-seq protocol. *Methods Mol Biol* 1708:83–104
12. Kernaleguen M, Daviaud C, Shen Y, Bonnet E, Renault V, Deleuze JF, Mauger F, Tost J (2018) Whole-genome bisulfite sequencing for the analysis of genome-wide DNA methylation and hydroxymethylation patterns at single-nucleotide resolution. *Methods Mol Biol* 1767:311–349
13. Dugé de Bernonville T, Maury S, Delaunay A, Daviaud C, Chaparro C, Tost J, O'Connor SE, Courdavault V (2020) Developmental methylome of the medicinal plant *catharanthus roseus* unravels the tissue-specific control of the monoterpene indole alkaloid pathway by DNA methylation. *Int J Mol Sci* 21:6028



14. Dugé de Bernonville T, Carqueijeiro I, Lanoue A, Lafontaine F, Sanchez Bel P, Liesecke F, Musset K, Oudin A, Glévarec G, Pichon O, Besseau S, Clastre M, St-Pierre B, Flors V, Maury S, Huguet E, O'Connor SE, Courdavault V (2017) Folivory elicits a strong defense reaction in *Catharanthus roseus*: meta bolomic and transcriptomic analyses reveal dis tinct local and systemic responses. *Sci Rep* 7: 40453
15. Goecks J, Nekrutenko A, Taylor J, Galaxy T (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and trans parent computational research in the life sciences. *Genome Biol* 11(8):R86
16. Xi Y, Li W (2009) BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinf* 10:232
17. Krueger F, Andrews SR (2011) Bismark: a flex ible aligner and methylation caller for Bisulfite Seq applications. *Bioinformatics* 27:1571–1572
18. Akalin A, Kormaksson M, Li S, Garrett Bakelman FE, Figueroa ME, Melnick A et al (2012) MethylKit: a comprehensive R package for the analysis of genome-wide DNA methyl ation profiles. *Genome Biol* 13:R87
19. Wu H, Xu T, Feng H, Chen L, Li B, Yao B, Qin Z, Jin P, Conneely KN (2015) Detection of differentially methylated regions from whole-genome bisulfite sequencing data with out replicates. *Nucleic Acids Res* 43(21):e141
20. Grehl C, Wagner M, Lemnian I, Glaser B, Grosse I (2020) Performance of mapping approaches for whole-genome bisulfite sequencing data in crop plants. *Front Plant Sci* 11:176