EpiGBS lab protocol

Niels Wagemaker (nwagemaker@science.ru.nl)
Radboud University Nijmegen
Philippine Vergeer
Wageningen University
Thomas van Gurp
NIOO
Koen Verhoeven
NIOO

Method Article

Keywords: EpiGBS, epigenetics, genomics, methylation, de novo, reference

Posted Date: February 15th, 2016

DOI: https://doi.org/10.1038/protex.2016.006

Abstract

Lab protocol of the Brief Communication, "epiGBS: a reference-free reduced representation bisulfite sequencing technique."

Introduction

We describe a reduced representation bisulfite sequencing method, named epiGBS, for cost-effective exploration and comparative analysis of DNA methylation and genetic variation in hundreds of samples of species de novo, without requiring a reference genome. This method uses genotyping-by-sequencing of bisulfite-converted DNA, followed by reliable de novo reference construction, mapping, variant calling and SNP/methylation variation distinction. The output can be loaded directly in IGV for visualization, and in RnBeads for analysis of differential methylation.

Reagents

DNA isolation kit : Macherey-Nagel nucleospin plant II Qubit® 2.0 Fluorometric dsDNA HS Assay Kit \(Q32851 Life technologies) PstI \(NEB, R0140S) BSA \(NEB, B9000S) T4 DNA ligase \(NEB, M0202M/L) non-phosphorylated adapters \(see supplementary data manuscript) Qiaquick PCR cleanup \(Qiagen, 28104) Agencourt AMPure XP \(Beckman coulter, A63880) 5-methylcytosine dNTP Mix \(Zymo research, D1030) DNA polymerase I \(NEB, M0209S) EZ DNA Methylation-Lightning™ Kit \(Zymo Research) 5 µL KAPA HiFi HotStart Uracil+ ReadyMix \(Kapa Biosystems) illumina PE PCR Primer \(see supplementary data manuscript) High Sensitivity DNA chip on a 2100 Bioanalyzer system \(Agilent) HiSeq v4 reagents

Equipment

Qubit® 2.0 Fluorometer Biometra Tgradient \(ramping speed 5 degrees / s) 2100 Bioanalyzer system \(Agilent) Illumina Hiseq 2500 sequencer; Rapid Run Mode Paired-End sequencing; HiSeq Control Software \(v2.2.38)

Procedure

A - DNA Extraction 1. Homogenise plant material by bead-beating frozen leaf tissue in a 2 mL eppendorf tube with 2-3 mm stainless steel beads. No more than 100 mg of fresh tissue was used per sample. 2. Take samples with beads from the freezer and store in liquid nitrogen. 3. Per batch of 12 samples put the tubes in a tissuelyzer 24 adapter set block, which is partially submerged in liquid nitrogen to avoid thawing during sample placement. 4. After 30 seconds of shaking at 30 hz / 1800 oscillations per minute, submerge the tubes in liquid nitrogen, then perform a second round of 30 second shaking. 5. Isolate DNA from the samples. For most plant species mentioned, we isolated DNA using the Macherey-Nagel nucleospin plant II kit \(manuscript : for details see Supplementary Excel file “Detailed sample description”) for individual tubes. We followed the manufacturer’s protocol with the following

modifications. Cell lysis was performed using Cell lysis buffer PL1 for 30 instead of 10 minutes. After filtration through the pink nucleospin filter the flow-through is carefully pipetted to a fresh 1.5 mL tube, avoiding the pellet, which is often formed at this stage. An additional centrifuge round was used to avoid a small pellet; the clear liquid was used in following steps according the manufacturer's protocol. As restriction enzymes are very sensitive to proteins and other contamination we only selected samples with high purity, specifically 260/280 and 260/230 ratios of at least 1.8 and 1.5 respectively). DNA concentration was determined using Qubit® 2.0 Fluorometric dsDNA HS Assay Kit \(Q32851 Life technologies). **B: PstI Restriction digestion** 6. Per individual, digest 400 ng of genomic DNA \(gDNA) overnight \(17hrs) at 37°C in a volume of 40 µL containing 1x NEBuffer 3.1, 125 µg BSA \(NEB, B9000S) and 2 µL / 40 units of PstI \(NEB, R0140S). Following digestion, ligate barcoded adapters to the fragments. \(for Csp6I alternative; see "Methods Csp6I Laboratory work") **C : Adapter ligation** To minimize the possibility of misidentifying samples as a result of sequencing or adapter synthesis error, all pair-wise combinations of barcodes differed by a minimum of three mutational steps, barcode lengths were modulated from 4 to 6 bp to maximize the nucleotide balance of the bases at each position in the overall set of sequencing reads \(manuscript : Supplementary Fig. 1d). Samples were pooled and processed per species after ligation. 7. For the ligation, combine 1200 pg of both forward and reverse barcoded adapters \(manuscript : Supplementary table Ab) in a 60 µL reaction containing 40 µL gDNA digest, 1x T4 DNA ligase buffer and 4000K units T4 DNA ligase \(NEB, M0202M/L) and ligate for 3hrs at 22°C followed by 4°C overnight, no inactivation afterward. **D: Cleanup and size selection** 8. Perform per-species pooling in order to assess the quality of libraries. 9. When pooled, reduce the total library volume by Qiaquick PCR cleanup \(Qiagen, 28104) to 60 µL. 10. Size select the libraries using a 0.8x Agencourt AMPure XP \(Beckman coulter, A63880) purification favoring >200 bp DNA fragments and elute in a total volume of 24 µL. **E: Nick translation** Due to the use of non-phosphorylated adapters, epiGBS libraries contain nicks between the 3' fragment overhang and the 5' non-phosphorylated adapter nucleotide. 11. To prevent the loss of ssDNA adapter strands \(at the nicked position) during bisulfite treatment, repair the nick \(see Supplementary Fig. 10) by a 1-hour nick translation reaction at 15°C in a reaction of 25 µL containing 18 µL of the purified library, 2,5 uL of 10 mM 5-methylcytosine dNTP Mix \(Zymo research, D1030), 1x NEBuffer 2 and 7,5 units DNA polymerase I \(NEB, M0209S). **Optional GBS PCR** At this stage an optional GBS PCR can be performed to check the library quality. 12. Perform GBS PCR using the epiGBS PCR protocol \(see below). The average size of the amplified GBS library was expected to be bigger than the amplified epiGBS library. 13. Assess the quality of these PCR-libraries by analyzing 1 µL on a High Sensitivity DNA chip on a 2100 Bioanalyzer system \(Agilent). **F: Bisulphite treatment and purification** For bisulfite treatment 20 µL of the nick-translated library was used. 14. Perform Bisulfite treatment using the EZ DNA Methylation-Lightning™ Kit \(Zymo Research) with the following program: 8 minutes 98°C, 1 hour at 54 °C followed by up to 20h at 4°C, all according to the manufacturer's protocol. **G: EpiGBS PCR** 15. Perform library amplification per species in four individual 10 µL reactions containing 1 µL ssDNA template, 5 µL KAPA HiFi HotStart Uracil+ ReadyMix \(Kapa Biosystems), 3 pmol of each illumina PE PCR Primer \(manuscript : Supplementary Table 1b). Temperature cycling consisted of 95°C for 3 min followed by 18 cycles of 98°C for 10 s, 65°C for 15 s, 72°C for 15 s with a final extension step at 72°C for 5 min. 16. Pool the replicate PCR products and

quantify using a Qubit® dsDNA HS Assay Kit \(Life Technologies). 17. Assess the quality of the Libraries analyzing 1 µL on a High Sensitivity DNA chip on a 2100 Bioanalyzer system \(Agilent). Libraries were considered suitable for sequencing if the majority of DNA fragments were between 150–400 bp and no adapter dimers were found. Typically, epiGBS PCR reactions of 18 cycles of a non-pooled plant sample yield 3-12 ng/µL of PCR-product. 18. When the 'per species' pooled libraries pass quality control they can be further pooled according to concentration and number of samples in the species pool so that each individual sample was expected to yield an equal number of clusters on the Illumina flowcell. 19. Perform A 'nano run' on the Illumina MiSeq to quantify per-sample expected relative read count yield. Based on the read counts obtained from this run, pool the individual nick-translated digestion-ligations in such a manner that an equal number of reads would be expected per individual. 20. Finally, perform Rapid Run Mode Paired-End sequencing on an Illumina HiSeq2500 sequencer using the HiSeq v4 reagents and the latest version of the HiSeq Control Software \(v2.2.38), which optimizes the sequencing of low-diversity libraries \(http://res.illumina.com/documents/products/technotes/technote-hiseq-low-diversity.pdf). As the first five cycles of a sequencing run are used to calculate the color matrix, our barcode design achieves almost perfect balance of the first 5 nucleotides when equal numbers of sequences are obtained per forward read or "A" barcode. The reverse read or "B" barcodes do not have this requirement, hence only barcodes of four nucleotides were used. **Methods Csp6I Laboratory work** Construct the Csp6I epiGBS libraries in similar fashion as the PstI epiGBS libraries with the following modifications: The restriction digestion reaction contained 1x FD buffer and 4 µL / 40 units of Csp6I \(ThermoFisher Scientific, FD0214). The ligation reaction contained 2400 pg of both A and B adapters \(both adjusted for the Csp6I sticky end). While in the PstI protocol we used fully methylated adapters \(both strand I and II methylated) for the Csp6I protocol we used hemi-methylated adapters. The adapter strands that were resynthesized \(incorporating 5mC dNTP's) by nick translation were not methylated as all cytosines are replaced by methylated 5mC \(manuscript : see Supplementary Fig. 10). Final library amplification for Csp6I yielded 4 - 8 ng / µL product for an epiGBS PCR of 18 cycles of a library only containing Arabidopsis sample A29.

Timing

Standard : 3 days excluding dna isolation Using optional Miseq nano sequencing before final pooling: 4 days excluding dna isolation

Troubleshooting

Essential for the protocol is perfect digestion of the gDNA. This can be inhibited by poor quality DNA and ethanol residues in gDNA sample. The best way to aproach this protocol is to start early in the week so the whole protocol can be performed in a 3 day sequence: day 1: digestion; day 2: ligation; day 3: pooling/purification, size selection, nick repair, BS conversion and purification, EpiGBS pcr. The overnight digestion / ligation might not be neccesary for all species. When pooling the adapter ligated reactions only a subsample is pooled \(amount dependent on number of samples). If perfect read distribution is

wanted an option miseq nano run can be performed with subsequent repooling according to individual read counts.

Anticipated Results

PCR will give sufficient product after 14 to 18 pcr cycles, depending on the amount of samples pooled and the species used. See also figure 1a \[step 8] of manuscript: Lab protocol for typical epiGBS library

Figures

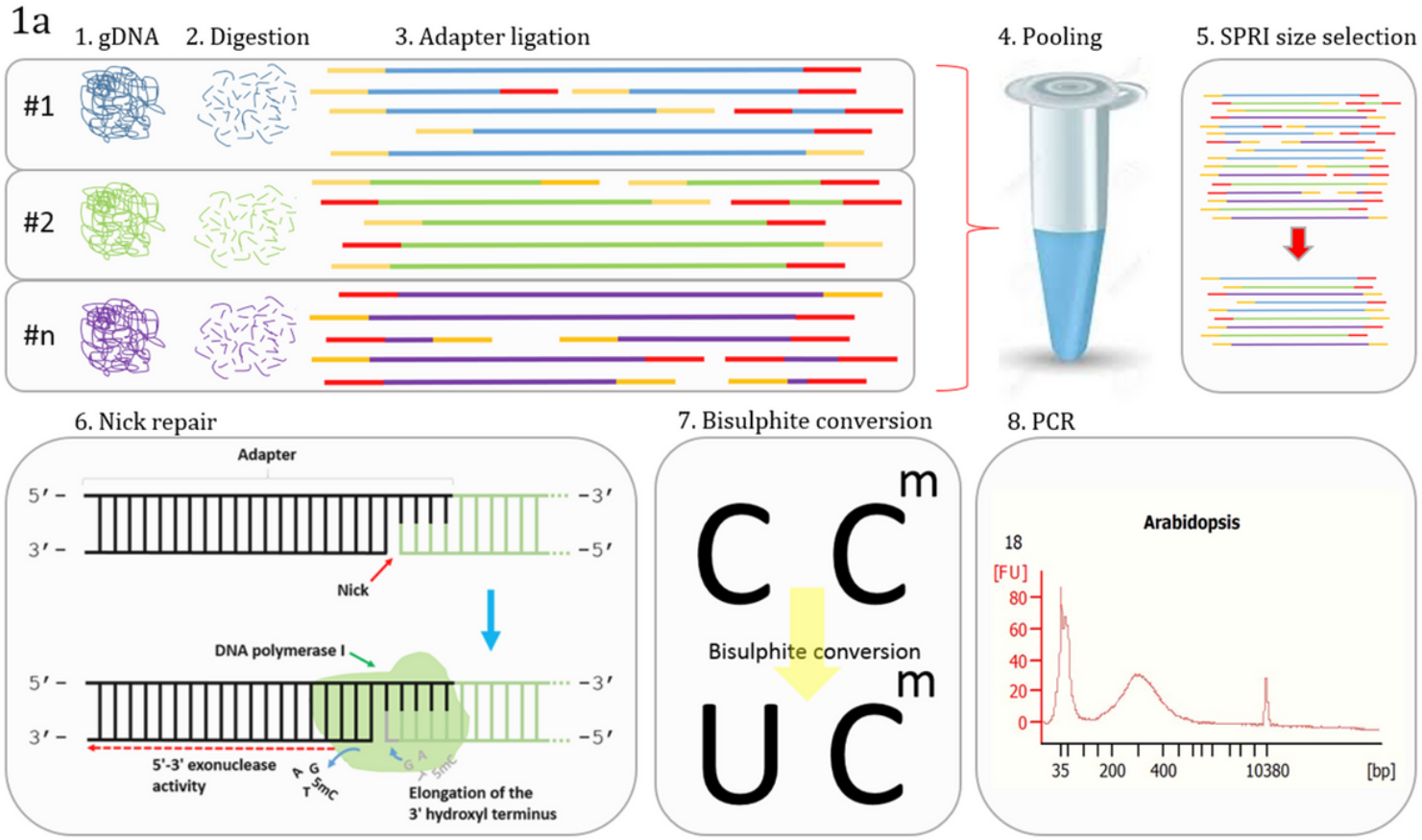

Figure 1

Lab protocol epiGBS lab protocol Genomic DNA (1) is digested with PstI (or other restriction enzyme of choice) (2) for sample 1-n. After barcoded adapter ligation (3) fragments are pooled (4), QIAquick PCR purified, and subjected to 0.8x SPRI size selection (5). Nick translation (6) is used to repair the nicks between the adapter and the restriction fragment. Fragments are bisulfite converted (7) and PCR amplified (8).