Accurate sodium bisulfite sequencing in plants

Ian R. Henderson,¹ Simon R. Chan,² Xiaofeng Cao,³ Lianna Johnson⁴ and Steven E. Jacobsen^{4,5,*}

¹Department of Plant Sciences; University of Cambridge; Downing Street, Cambridge, UK; ²Section of Plant Biology; University of California, Davis; Davis, CA USA; ³State Key Laboratory of Plant Genomics and National Center for Plant Gene Research; Institute of Genetics and Developmental Biology; Chinese Academy of Sciences; Beijing, China; ⁴Department of Molecular, Cellular and Developmental Biology; University of California, Los Angeles; Los Angeles, CA USA; ⁵Howard Hughes Medical Institute; University of California, Los Angeles; CA USA

Key words: DNA methylation, plants, bisulfite, silencing

DNA cytosine methylation is a conserved epigenetic modification frequently correlating with transcriptional silencing in a wide variety of eukaryotic organisms. Sodium bisulfite treatment of DNA converts unmethylated cytosine to uracil, while 5-methylated cytosine is protected. We describe techniques that ensure reliable sequencing data following sodium bisulfite conversion and to avoid common pitfalls such as amplification of unconverted DNA and inclusion of sibling clones.

Cytosine methylation is commonly found on repeated sequences and silent loci, though it is also observed on expressed genes.¹⁻⁴ Plants display cytosine methylation in CG, CHG and CHH (where H is any nucleotide apart from guanine) sequence contexts. Understanding the function of this epigenetic mark requires techniques to accurately assess its distribution. A useful method to analyze cytosine methylation is sodium bisulfite sequencing.^{5,6} Treatment of DNA with sodium bisulfite causes deamination of cytosine to uracil, unless this reaction is blocked by methylation at the 5-carbon position.^{5,6} Amplification of bisulfite treated DNA by polymerase chain reaction (PCR) leads to uracil being amplified as thymine, whereas methylated cytosine remains as cytosine.^{5,6} Sequencing of the amplified DNA is then used to score the frequency with which sites are present as either cytosine or thymine.^{5,6} This serves as a measure of methyl-cytosine frequency in the original DNA sample. Sequencing can be performed following amplification and cloning of specific genomic regions into bacterial vectors.^{5,6} Recent advances in high-throughput sequencing have also been combined with bisulfite conversion, to analyze DNA methylation patterns on a genome-wide scale.^{1,2} The main advantage of these techniques is that they provide single base-pair resolution of methylation patterns. In plants this is particularly useful as cytosine sequence context can be determined, which can have important implications for the mechanism of methylation maintenance through cell division.7

Sodium bisulfite sequencing is a reliable technique when employed carefully but is prone to a number of artifacts, especially when applied to plant systems, which can show methylation in any sequence context. Here we draw attention to potential pitfalls and describe simple techniques to avoid them. A common problem in sodium bisulfite sequencing is amplification of unconverted genomic DNA. After sequencing this is evident as clones with strings of many adjacent "methylated" cytosines in all sequence contexts (Fig. 1A). Genome-wide analysis of cytosine methylation in *Arabidopsis thaliana* has shown that CHG and CHH sites are on average methylated at 6.7 and 1.7%, and that the methylation status of adjacent sites do not show a high correlation in most instances.^{1,2} Hence, observation of long stretches of adjacent methylated sites almost always indicates amplification of unconverted DNA (Fig. 1A). In our experience, more stringent bisulfite conversion protocols eliminate this artifact. Incomplete denaturation of the template DNA contributes greatly to this problem. It is of course conceivable that very high levels of methylation in all sequence contexts are truly found at some loci. In this instance results should be verified using alternative techniques that do not use a bisulfite conversion step. For example, Southern blotting combined with digestion using methyl-sensitive restriction endonucleases.⁸

A key step to reduce the likelihood of amplifying unconverted DNA is to design primers biased to amplify fully converted DNA. The average length of DNA fragments present after conversion will vary according to protocol and whether the sample was treated enzymatically, for example by restriction digestion. As sodium bisulfite treatment is damaging to the template DNA it is typically difficult to amplify products greater than 500 base pairs from converted DNA; so a region shorter than this should be selected for study to avoid extreme bias toward longer unconverted (and undamaged) fragments. A single primer pair allows analysis of one DNA strand, though hairpin-bisulfite strategies allow both strands to be analyzed simultaneously.⁹

As unmethylated cytosines will be converted to uracil it is important to choose a relatively G-rich region when designing the top-strand primer. This ensures that a sufficiently high annealing temperature can be used without an excessively long oligonucleotide. All cytosines in the primer should be changed to thymine, with the exception of the generally highly methylated CG sites,

Submitted: 10/13/09; Accepted: 11/08/09

^{*}Correspondence to: Steven E. Jacobsen; Email: jacobsen@ucla.edu

Previously published online: www.landesbioscience.com/journals/epigenetics/article/10560



Figure 1. Analysis of cytosine methylation by amplification and sequencing following sodium bisulfite conversion of genomic DNA. (A) Hypothetical data illustrating the difference between sequencing reads generated from fully bisulfte converted and unconverted DNA samples. The original genomic sequence is indicated below with ten hypothetical sequencing reads stacked above. CG sites are highlighted in red, CHG sites in orange and CHH sites in blue. Unmethylated sites are evident in the sequencing reads when a cytosine (C) is replaced by thymine (T). (B) Design of primers to amplify sodium bisulfite converted DNA. The genomic sequence is shown above the corresponding bisulfite sequencing primer with the changes highlighted. (C) Hypothetical data illustrating the presence of sibling clones within the sequencing reads.

which should be changed to Y (C or T) (Fig. 1B). As CG sites are more frequently methylated, even fully converted DNA is likely to remain as a C and using Y in primers increases the likelihood that the primer can hybridize effectively. Where possible the number of CG sites in the primer should be minimized (fewer than three is ideal) to reduce degeneracy. The primer should terminate at one or multiple cytosines in the CHH sequence context (Fig. 1B). As most CHH sites are methylated at a low frequency this will create a primer with a 3'-end showing a hybridization preference for fully converted DNA. One caveat is that this will create a bias to amplify DNA molecules that are unmethylated at these specific CHH sites. However, comparison of sequencing data generated using primers designed this way with independent methods does not reveal significant differences.^{1,2} Furthermore, one can use available data on preferred sequence contexts for CHH methylation to further select for low methylation sites.¹ The length of the primer should be adjusted such that its annealing temperature is above 65°C, which means primer length may become long (>30 nucleotides). Similar rules are applied to the bottom strand. In this case we select a C-rich region (on the bottom strand) and convert all guanine to adenine, with the exception of CG sites, which should be changed to R (G or A) (Fig. 1B). Again, we follow the same rules with respect to the 3'-end and length of the primer (Fig. 1B).

During PCR amplification primers are added using 55°C as the hybridization temperature and 60°C as the elongation temperature, for 40 cycles. After the amplification reaction is complete it should be analyzed using gel electrophoresis to confirm that the expected size of PCR product has been obtained. Gel purification is recommended to remove any primer dimers. The purified PCR amplification product can then be cloned and sequenced using conventional methods.

A second critical consideration is that each sequencing trace should represent an independent DNA molecule. Sibling clones can be recognized during sequencing as clones with identical patterns of methylation. As CHH sites are typically methylated at low frequency, the chances of two independent clones possessing an identical CHH methylation distribution are very unlikely.¹

References

- Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, et al. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature 2008; 452:215-9.
- Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, et al. Highly integrated singlebase resolution maps of the epigenome in Arabidopsis. Cell 2008; 133:523-36.
- Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, et al. Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. Cell 2006; 126:1189-201.
- Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S. Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. Nat Genet 2007; 39:61-9.
- Clark SJ, Harrison J, Paul CL, Frommer M. High sensitivity mapping of methylated cytosines. Nucleic Acids Res 1994; 22:2990-7.
- Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci USA 1992; 89:1827-31.

©2010 Landes Bioscience. Do not distribute.

Hence, if one or more clones show identical CHH patterns, only one should be included for analysis to reduce the chance that the same DNA molecule is being re-analyzed (Fig. 1C). Amplification of sibling clones is frequently a problem when using nested PCR amplifications, a large number of amplification cycles or low amounts of starting DNA.

Application of these simple rules and awareness of potential problems should enable the generation of accurate and reproducible sodium bisulfite sequencing data to analyze patterns of cytosine DNA methylation in plants.

- Henderson IR, Jacobsen SE. Epigenetic inheritance in plants. Nature 2007; 447:418-24.
- McClelland M, Nelson M, Raschke E. Effect of sitespecific modification on restriction endonucleases and DNA modification methyltransferases. Nucleic Acids Res 1994; 22:3640-59.
- Laird CD, Pleasant ND, Clark AD, Sneeden JL, Hassan KM, Manley NC, et al. Hairpin-bisulfite PCR: assessing epigenetic methylation patterns on complementary strands of individual DNA molecules. Proc Natl Acad Sci USA 2004; 101:204-9.