Bioinformatics, 34(4), 2018, 708–709
doi: 10.1093/bioinformatics/btx633
Advance Access Publication Date: 26 October 2017
Applications Note

Data and text mining

ViewBS: a powerful toolkit for visualization of high-throughput bisulfite sequencing data

Xiaosan Huang[1], Shaoling Zhang[1], Kongqing Li[2], Jyothi Thimmapuram[3] and Shaojun Xie[3,*]

[1]College of Horticulture, State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, No.6 Tongwei Road Nanjing, P. R. China, 210095, [2]Department of Rural Development, Nanjing Agricultural University, No.6 Tongwei Road Nanjing, P. R. China, 210095 and [3]Bioinformatics Core, Purdue University, 155 South Grant Street, West Lafayette, Indiana 47907, USA

*To whom correspondence should be addressed
Associate Editor: Jonathan Wren

Received on June 28, 2017; revised on September 14, 2017; editorial decision on October 1, 2017; accepted on October 26, 2017

Abstract

Motivation: High throughput bisulfite sequencing (BS-seq) is an important technology to generate single-base DNA methylomes in both plants and animals. In order to accelerate the data analysis of BS-seq data, toolkits for visualization are required.
Results: ViewBS, an open-source toolkit, can extract and visualize the DNA methylome data easily and with flexibility. By using Tabix, ViewBS can visualize BS-seq for large datasets quickly. ViewBS can generate publication-quality figures, such as meta-plots, heat maps and violin-boxplots, which can help users to answer biological questions. We illustrate its application using BS-seq data from *Arabidopsis thaliana*.
Availability: ViewBS is freely available at: https://github.com/xie186/ViewBS.
Contact: xie186@purdue.edu
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

A combination of bisulfite treatment of DNA and high-throughput sequencing [bisulfite sequencing (BS-seq)] has been extensively used to investigate DNA methylation at single-base resolution level for animals and plants (Krueger *et al.*, 2012). Analysis of BS-seq data includes alignments of reads, identification of differentially methylated regions (DMR) and visualization of the data. Multiple tools for read alignments of BS-seq data have been developed (Guo *et al.*, 2013; Harris *et al.*, 2012; Krueger and Andrews, 2011; Pedersen *et al.*, 2014). Many tools (Akalin *et al.*, 2012; Hansen *et al.*, 2012; Wang *et al.*, 2015) are available for DMR identification as well. To visualize the DNA methylome data, heat map, meta-line plot and boxplot are frequently used in published papers (Krueger *et al.*, 2012). Developing a toolkit with functions to generate publication-ready figures will be broadly used in the analysis of BS-seq data.

Here we have developed ViewBS—a standalone program capable of both profiling genome-wide DNA methylation and visualizing DNA methylation patterns of BS-seq data at selected regions. We used a few examples to demonstrate that ViewBS is easy to use and very powerful to generate publication-quality figures.

2 Implementation

ViewBS has one main command named ViewBS. Under ViewBS, eight tools are developed. Getopt:: Long:: Subcommand is used to process command line options with sub commands. Bio:: DB:: HTS:: Tabix (another Perl module) is used to quickly retrieve genome-wide cytosine report as input. Perl module Bio:: SeqIO is used to retrieve information from sequence data. To generate publication-quality figures, three packages (reshape2, ggplot2 and pheatmap) are used in R. The source code is freely available at https://github.com/xie186/ViewBS.

3 Functions and examples

ViewBS has several tools which determine the required and optional arguments. These tools can be divided into two parts: profiling of genome-wide DNA methylation and visualization of DNA

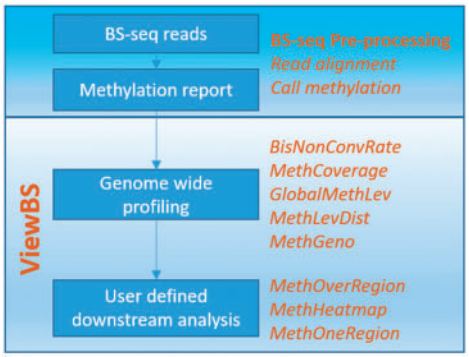

Fig. 1. Summary of the functions of ViewBS

methylation patterns of BS-seq data at selected regions (Fig. 1). Each tool will generate the result in tab-delimited files, corresponding figures in PDF files and shell scripts to regenerate the figures.

To use ViewBS, users typically need to prepare two types of data set. ViewBS uses genome-wide cytosine methylation report generated by Bismark as input file, which contains the sequence context and has seven columns in the following format: chromosome, position, strand, count methylated, count unmethylated, C-context and trinucleotide context. The second one is input file for selected regions of interest which includes the genomic coordinates.

To evaluate the performance of ViewBS, we used the BS-seq data (see Supplementary Material) for *Arabidopsis thaliana* (Stroud *et al.*, 2014). Figures generated by the tools of ViewBS were based on these data. Details can be found in supplementary data.

Profiling of genome-wide DNA methylation

For genome-wide DNA methylation profiling, we offer several tools: *BisNonConvRate*, *MethCoverage*, *GlobalMethLev*, *MethGeno* and *MethLevDist*.

BisNonConvRate is a tool for estimating non-conversion rate of BS-seq data. In plant, reads mapping to the non-methylated chloroplast genome can be used to assess bisulfite conversion efficiency. BisNonConvRate is especially useful in this case. Users can provide the chromosome ID for chloroplast and *BisNonConvRate* will estimate the non-conversion rate.

MethCoverage can be used to assess the read coverage distribution of cytosine in each context. This tool is useful to let users know the read coverage of BS-seq data.

Global DNA methylation levels are common information for understanding the BS-seq data. *GlobalMethLev* is the tool to generate weighted DNA methylation levels for the samples that the users provide.

Distribution of methylation levels is another feature that researchers use to profile DNA methylation data. The tool *MethLevDist* can generate methylation level distributions for not only genome-wide, but also for the regions of interest, for example genic regions, TSS regions, etc.

Another tool named *MethGeno* can generate the methylation levels across the chromosomes.

Visualization of DNA methylation patterns at selected regions of interest

For visualization of DNA methylation patterns at functionally important regions, we offer three tools: *MethOverRegion*, *MethHeatmap* and *MethOneRegion*.

MethOverRegion is a tool which can generate meta-plot for selected regions, for example a list of genes. By default, the selected region will be split into 60 bins. The flanking 2 kb regions of the selected regions will be split into 100-bp bins. For each bin, weighted methylation levels will be recorded and plotted along the selected regions. The users can set the size of flanking regions and the number of bins.

MethHeatmap is a tool that can be used to generate heat map and/or violin-box plot for selected regions, for example a list of DMRs. A violin plot shows the full distribution of the data (Hintze and Nelson, 1998) and a box plot shows summary statistics such as mean/median and interquartile ranges. Here *MethHeatmap* combines these two methods together into one violin-boxplot to visualize the data.

MethOneRegion is a tool which can be used to visualize DNA methylation information for just one region provided. This is useful if the users don't want to load the large dataset to genome browser (like IGV Thorvaldsdottir et al., 2013) and just want a snapshot of one region. In this tool, the users can define the size of flanking regions that they want to visualize. The DNA methylation in the region provided will be shown in shaded transparent area.

Evaluation of ViewBS was carried out using BS-seq data from *Arabidopsis*. Details of the results, including the maximum memory consumptions and time used for each of the tools of ViewBS, were shown in the supplementary files.

4 Conclusions

We conclude that ViewBS is an extremely efficient and flexible software package to accelerate research in the era of bisulfite sequencing data. It provides a set of toolkits to enable rapid analysis of whole genome bisulfite sequencing.

Acknowledgements

We thank Dr Hainan Zhao for the suggestions on the project.

Funding

This work has been supported by the Excellent Youth Natural Science Foundation of Jiangsu Province (SBK2017030026), the Natural Science Foundation of Jiangsu Province (BK20150681), the National Natural Science Foundation of China (71704081), the Fundamental Research Funds for the Central Universities (KYTZ201401, SKYC2017019, KYZ201607, KYZ201510 and SKTS2017015), the Ministry of education of Humanities and Social Science project (14YJC630058).

Conflict of Interest: none declared.

References

Akalin,A. *et al.* (2012) methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol*, **13**, R87.

Guo,W. *et al.* (2013) BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. *BMC Genomics*, **14**, 774.

Hansen,K.D. *et al.* (2012) BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol*, **13**, R83.

Harris,E.Y. *et al.* (2012) BRAT-BW: efficient and accurate mapping of bisulfite-treated reads. *Bioinformatics*, **28**, 1795–1796.

Hintze,J.L., and Nelson,R.D. (1998) Violin plots: a box plot-density trace synergism. *Am. Stat*, **52**, 181–184.

Krueger,F. *et al.* (2012) DNA methylome analysis using short bisulfite sequencing data. *Nat Methods*, **9**, 145–151.

Krueger,F., and Andrews,S.R. (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, **27**, 1571–1572.

Pedersen,B.S. *et al.* (2014) Fast and accurate alignment of long bisulfite-seq reads. https://arxiv.org/abs/1401.1129.

Stroud,H. *et al.* (2014) Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*. *Nat. Struct. Mol. Biol*, **21**, 64–72.

Thorvaldsdottir,H. *et al.* (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinformatics*, **14**, 178–192.

Wang,Z. *et al.* (2015) swDMR: a sliding window approach to identify differentially methylated regions based on whole genome bisulfite sequencing. *PLoS One*, **10**, e0132866.