Comparison of Bioinformatics Pipelines to Analyze Fish Methylomes in Relation to Environmental Adaptation

Mirella Arguelles[1], Keem Uarren Eiryll B. Antonio[2*], Mylah Rystie U. Anacleto[1], and Jorge Gil C. Angeles[2]

[1]Institute of Computer Science, College of Arts and Sciences [2]Philippine Genome Center - Program for Agriculture, Livestock, Forestry, and Fisheries, University of the Philippines Los Baños
*Corresponding author: kbantonio1@up.edu.ph

ABSTRACT

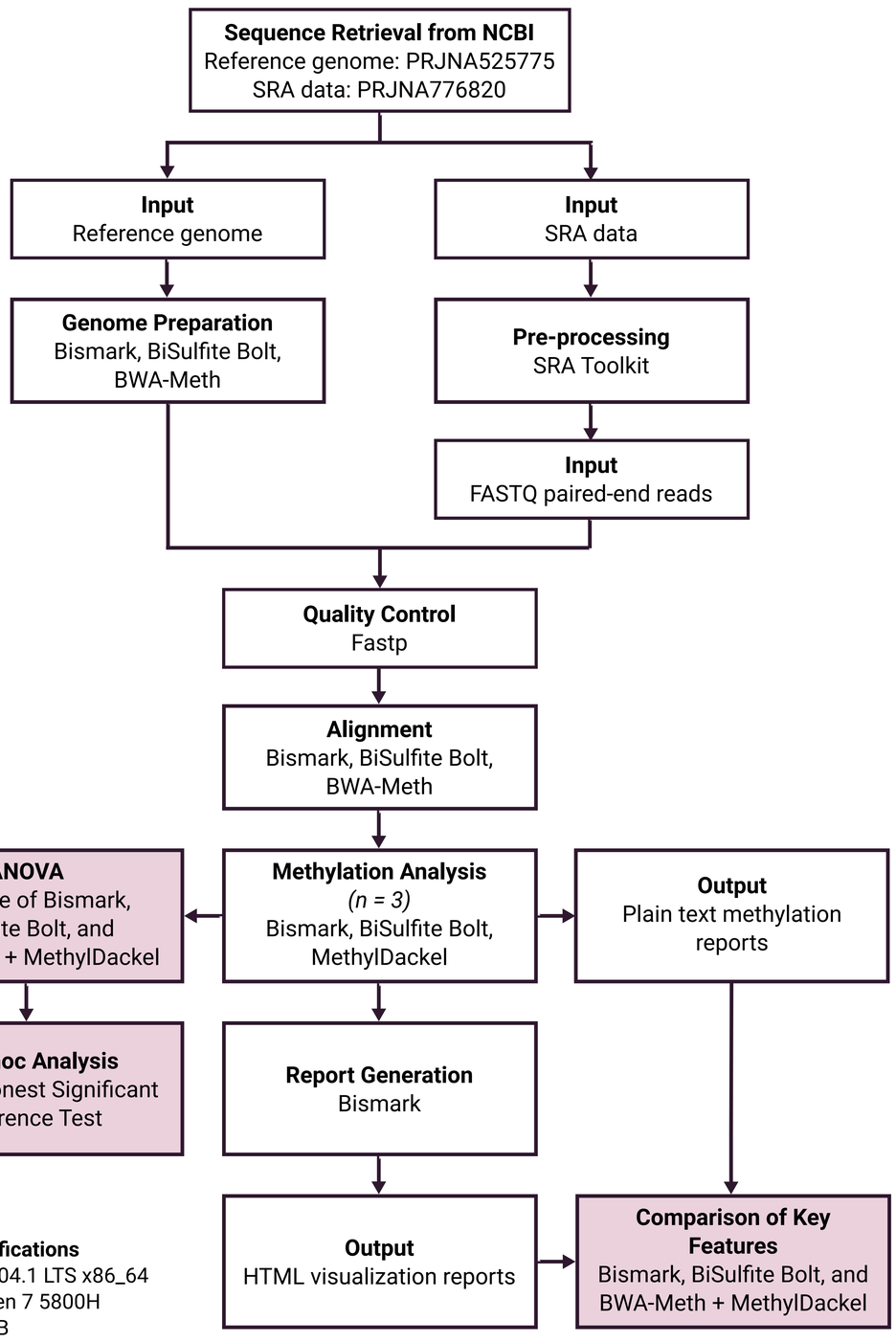
Methylation of DNA is a well-studied epigenetic mechanism and plays an important role in gene regulation and expression of organisms. Through the recent advancements of next-generation sequencing (NGS) technologies, generation of high-throughput methylation data, and elucidation of DNA methylation patterns; insights on how DNA methylation affects the organism's adaptation, such as with fishes, to its environment could be attained. Recognition of methylation patterns through NGS and the subsequent detection of significant methylation differences are some challenges to effectively analyze fish methylomes. There are several bioinformatics pipelines employed to analyze methylomes in mammals, insects, and plants but none have been comparatively used to analyze fish methylomes especially to understand their environmental adaptation. In this study, three commonly-used methylation bioinformatic pipelines (Bismark, BiSulfite Bolt, and BWA-Meth) were tested on the sole available methylome data relating to the environmental adaptation of fishes, that from the three-spined stickleback (*Gasterosteus aculeatus*). Comparative analyses focusing on the performance and functionality of these three programs were conducted to detect each program's advantages and limitations. Bismark and BiSulfite Bolt have no significant differences relating to their run time and are significantly faster than BWA-Meth to analyze the given fish methylome data. Comparing the primary features of the three programs, Bismark has the further advantage to remove duplicated reads given different biological samples. It also has features that would greatly improve data visualization and provide summary reports that can be used for downstream analyses. The findings of this study could be adapted for future methylome analyses concerning the environmental adaptation of fishes.

INTRODUCTION

Epigenetics is the study of changes in gene expression that do not involve alterations to the DNA sequence. DNA methylation involves the addition of a methyl group at the 5th carbon of cytosine. It is the most widely studied epigenetic mark and affects a variety of biological processes of an organism such as its development and adaptation to its environment. Through next-generation sequencing techniques, bisulfite modification is employed to study changes of DNA methylation patterns. Several bioinformatics pipelines which specialize in understanding bisulfite sequencing data have been used to identify such methylation patterns. These pipelines have been used in various literature to determine which is most suited for mammals, insects, and plants. However, none have been studied for fish methylomes especially to understand their environmental adaptation.

This study conducts a comparative analysis on commonly used methylome analysis pipelines using available methylomic reads from the three-spined stickleback (*Gasterosteus aculeatus*). The comparison between the tools were based on run time and evaluation of their primary features. Identifying a suitable pipeline for the project may improve the accuracy and efficiency of DNA methylation analysis for fish methylomes. The results provide researchers with adequate direction on which methylation analysis pipeline to choose for their studies.

METHODOLOGY



Machine Specifications
OS: Ubuntu 22.04.1 LTS x86_64
CPU: AMD Ryzen 7 5800H
Memory: 512GB
Number of threads used: 4

RESULTS

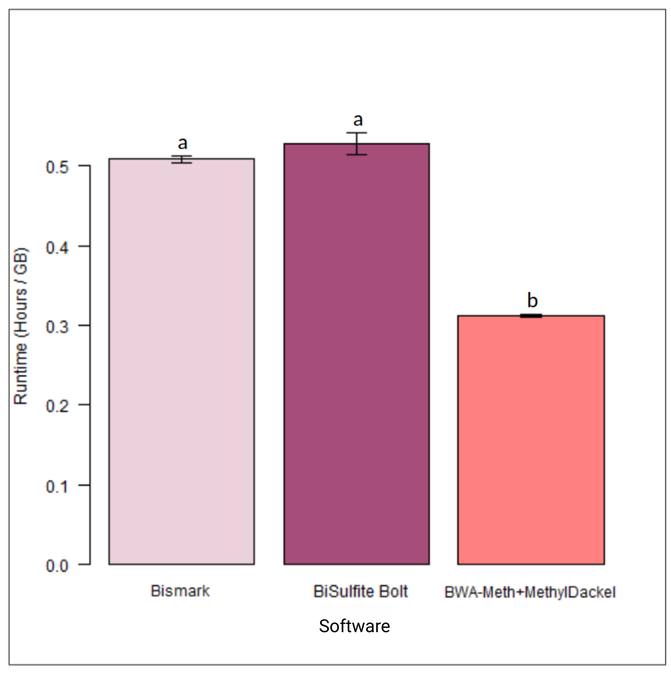


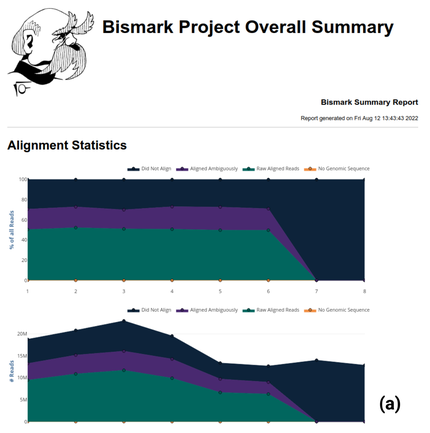
Figure 1. Normalized run times (n = 3) of three different methylation analysis pipelines using available methylome data relating to the environmental adaptation of three-spined stickleback (*Gasterosteus aculeatus*). Software with the same letter have no statistical difference (α < 0.01).

- **No significant differences** between Bismark and BiSulfite Bolt (α < 0.01)
- Both Bismark and BiSulfite Bolt have **faster run time** than BWA-Meth (α < 0.01)

Table 1. Summary of key features of three different methylation analysis pipelines

Feature	Bismark	BiSulfite Bolt	BWA-Meth and MethylDackel
Reference genome indexing	✓	✓	✓
Alignment of bisulfite sequencing reads	✓	✓	✓
Deduplication of sequencing reads	✓	✗	✗
Methylation calling	✓	✓	✓
Processing report generation	✓	✗	✓
Summary report generation	✓	✗	✗
HTML report visualization	✓	✗	✗
Python API availability	✗	✓	✗
Standalone software	✓	✓	✗

- Bismark has features for **deduplication of sequencing reads** and **plot visualization** of reports while these features do not exist in BiSulfite Bolt and BWA-Meth.
- BWA-Meth is **not** a **standalone** methylation analysis software as it requires MethylDackel to perform methylation calls and generate analysis reports.



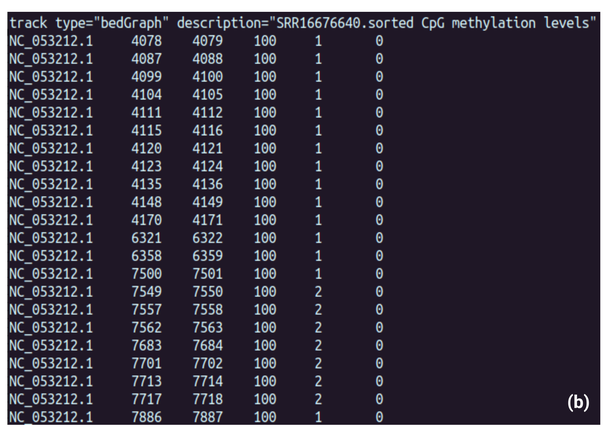


Figure 2. Methylation report outputs of (a) Bismark and (b) BiSulfite Bolt

CONCLUSION

From the benchmarking tests performed using fish methylome data, it was found out that Bismark and BiSulfite Bolt had the shortest time performance and both were faster than BWA-Meth. Bismark had more comprehensive features in comparison to BiSulfite Bolt and BWA-Meth. The results show that Bismark is a suitable candidate to analyze fish methylomes and other methylomes relating to environmental adaptation.

SELECTED REFERENCES

Artemov, A. V., et al. (2017). Genome-Wide DNA Methylation Profiling Reveals Epigenetic Adaptation of Stickleback to Marine and Freshwater Conditions. *Molecular Biology and Evolution, 34*(9), 2203–2213. https://doi.org/10.1093/molbev/msx156
Gong, W., et al. (2022). Benchmarking DNA methylation analysis of 14 alignment algorithms for whole genome bisulfite sequencing in mammals. *Computational and Structural Biotechnology Journal, 20*, 4704–4716. https://doi.org/10.1016/j.csbj.2022.08.051
Farrell, C., et al. (2021). BiSulfite Bolt: A bisulfite sequencing analysis platform. *GigaScience, 10*(5), giab033. https://doi.org/10.1093/gigascience/giab033
Krueger, F., & Andrews, S. R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics (Oxford, England), 27*(11), 1571–1572. https://doi.org/10.1093/bioinformatics/btr167