Benchmarking of sequencing technologies for CpG methylation detection: applications to quail and pig

Céline Vandecasteele, Paul Terzian, Joanna Lledo, Rémy-Félix Serre, Jules Sabban, Claire Kuchly, Frédérique Pitel, Sophie Leroux, Julie Demars, Nathalie Iannuccelli, et al.

Benchmarking of sequencing technologies for CpG methylation detection: applications to quail and pig

Céline Vandecasteele[1,*], Paul Terzian[2,*], Joanna Lledo[1], Rémy-Félix Serre[1], Jules Sabban[1], Claire Kuchly[1], Frédérique Pitel[3], Sophie Leroux[3], Julie Demars[3], Nathalie Iannuccelli[3], Christine Gaspin[2], Denis Milan[1,3], Carole Iampietro[1], Christophe Klopp[2] and Cécile Donnadieu[1]

[1]INRAE, US 1426, GeT-PlaGe, Genotoul, France Genomique, Université Fédérale de Toulouse, Castanet-Tolosan, France. [2]Université Fédérale de Toulouse, INRAE, BioinfOmics, GenoToul Bioinformatics facility, 31326, Castanet-Tolosan, France. [3]GenPhySE, Université de Toulouse, INRAE, INPT, ENVT, Castanet-Tolosan Cedex, F-31326, France. [4]Université Fédérale de Toulouse, INRAE, MIAT, 31326, Castanet-Tolosan, France.
*These authors contributed equally to this work.

BACKGROUND & SUMMARY

Context : Current technologies for detecting and analyzing epigenetic modifications are not yet as advanced or as cost-effective as those used for genetic analysis. This makes large-scale screening of epigenetic traits in breeding populations challenging and expensive. Advances in technology and a deeper understanding of epigenetic mechanisms will likely enhance the effectiveness and precision of epigenetic breeding in the future.

Objectives : Provide research teams with a powerful and innovative method to study methylation marks in a large number of individuals within breeding programs.

Questions : Is Whole Genome Bisulfite Sequencing (WGBS) still the gold standard for DNA methylation analysis ? Is Enzymatic Methyl-seq (EM-seq) a better reference to represent the short-read sequencing ? Technologies like Oxford Nanopore (ONT) and PacBio offer the advantage of long reads but are they powerful for detecting methylation ?

METHODS

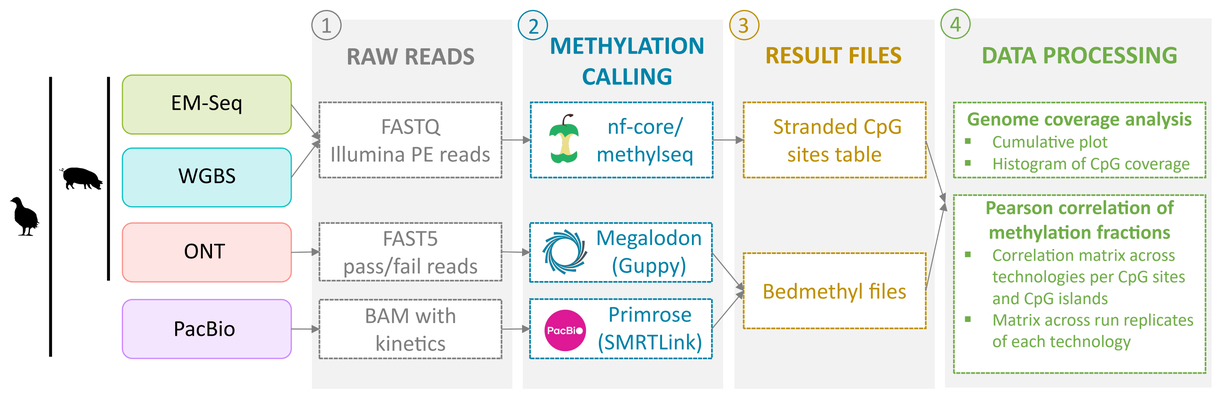

Figure 1. Overview of study design and dataset produced.
The quail and pig samples were sequenced using EM-Seq and WGBS methods on Illumina HiSeq3000 or NovaSeq6000, as well as Oxford Nanopore technology on PromethION from R9.4.1 flow cell. Additionally, the quail sample was sequenced using PacBio on Sequel2. An average coverage of 30x was achieved for each method. The dataset was then processed by 1) analyzing the format of raw reads produced for each method, 2) applying specific CpG methylation calling pipelines, 3) compiling fractions of 5mC methylation and depth of coverage per CpG position and strand (one file per run), and 4) performing R analysis with result data to evaluate and compare the technologies.

RESULTS

Comparing CpG sites coverage across technologies :

At the 10X threshold, the WGBS approach covers less than 75% of CpG sites, while EM-Seq and PacBio cover around 90% and ONT achieves close to 100% coverage.
WGBS shows significant heterogeneity in CpG site coverage, along with a high proportion of CpG sites uncovered, indicating a substantial alignment bias along the genome.

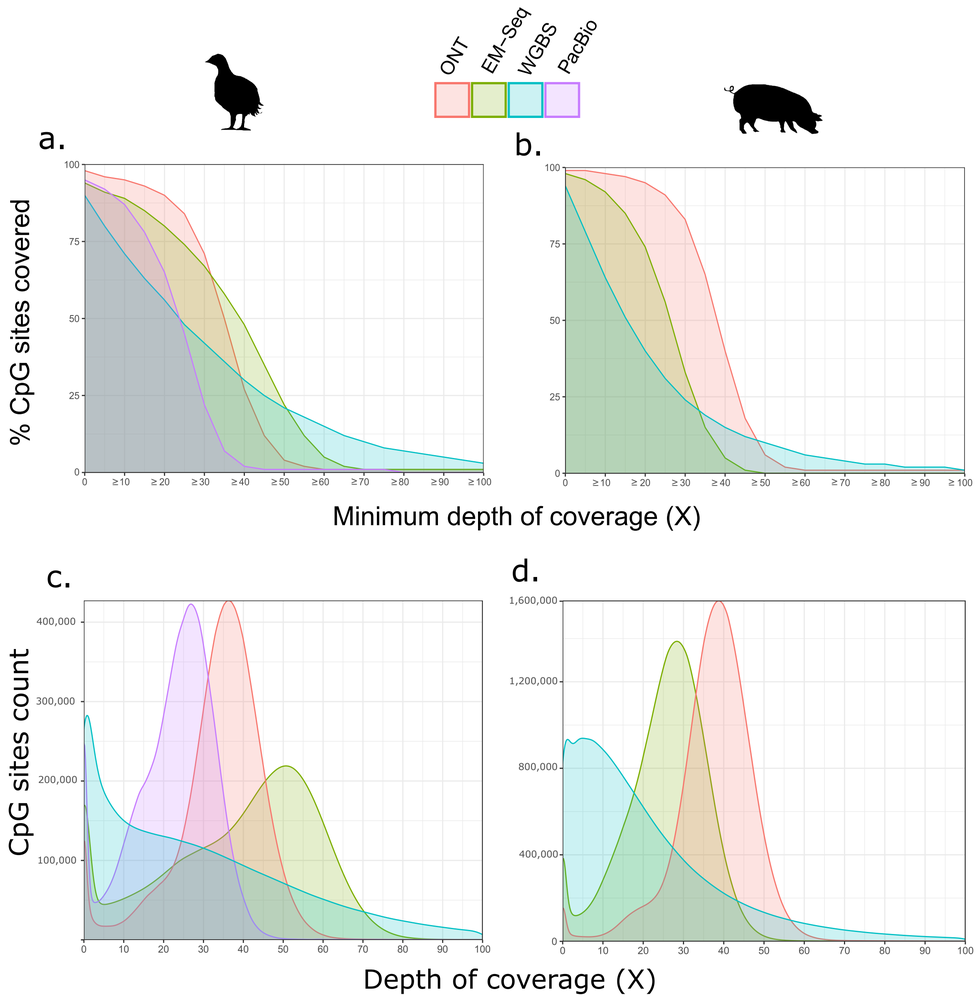

Figure 2. Coverage of CpG sites across technologies.
The percentage of CpG sites covered at a minimum depth of coverage (a,b) and the distribution of CpG site coverage (c,d) for ONT (red), EM-seq (green), WGBS (blue) and PacBio (purple) sequencing of *Coturnix japonica* (a,c) and *Sus scrofa* (b,d) samples.

Comparing CpG methylation measurements across technologies :

The highest correlation is always obtained between ONT and EM-Seq.

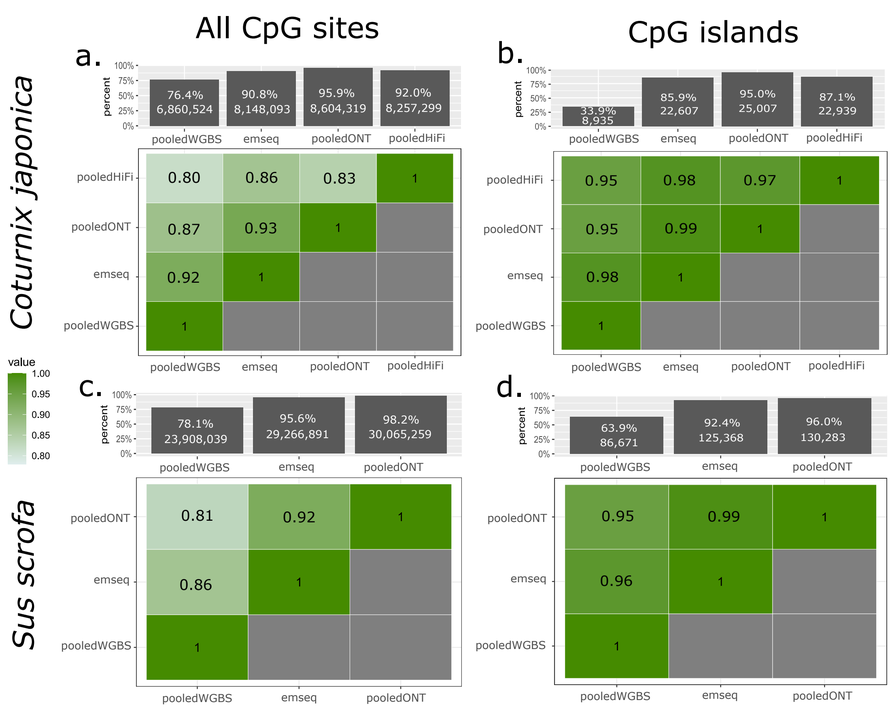

Figure 3. Pearson correlation matrix of CpG methylation level per-site and island across technologies.
Correlation analysis using (a,c) all shared CpG positions and (b,d) the average methylation fractions in CpG islands. Only CpG islands where at least 40% of CpG sites covered by ten or more reads were included in the correlation analysis. Barplots above each matrix show the number of CpG islands that passed this criteria (b,d), the number of CpG sites covered by 5 or more reads (a,c) and their percentage compared to the reference genome for each technology. Panels (a) and (b) report results for quail DNA, whereas (b) and (d) detail findings for pig DNA.

ONT released R10.4.1 flow cell methylation calling enhancements :

R10.4.1 flow cell outperformed the R9.4.1 flow cell in methylation calling using the remora super-accurate model.

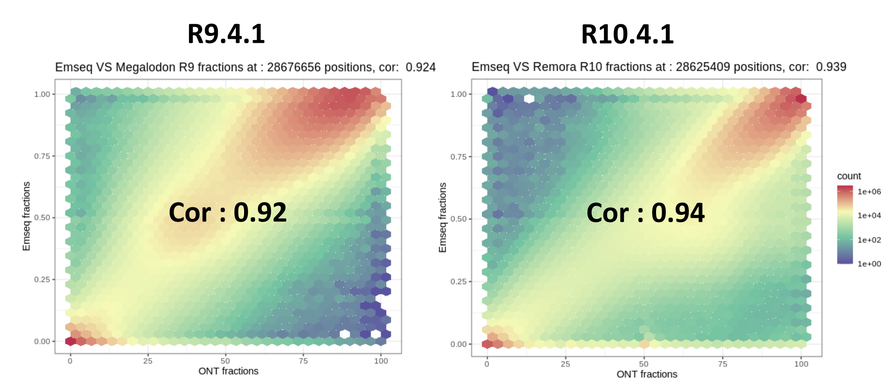

Figure 4. Correlation heatmap between Em-Seq versus ONT R9.4.1 and R10.4.1 flow cells
The 5mC DNA methylations in CpG site were detected by Megalodon v.2.5.0 with guppy v5.0.17 using res_dna_r941_prom_modbases_5mC_CpG_v001.cfg as calling model for R9.4.1 flow cell and Remora super-accurate model for R10.4.1 flow cell. Both were compared to the same Em-Seq data from pig sample.

CONCLUSION

Choosing the most suitable technology often involves balancing resolution, genome coverage, sample throughput, technical complexity, and overall cost. Our study [1] and other published benchmarks [2-6] demonstrate the performance of Nanopore Technology for methylation detection in addition to its advantage of long reads, which are particularly useful for complex genomic regions or for detecting structural variations alongside methylation. For this reason, the SeqOccIn project continued to use Nanopore technology on a larger number of samples in pig and quail, allowing the study of intergenerational epigenetic inheritance [7,8].
However, ONT is still more expensive than short-read technologies and poses challenges for use in breeding programs. For projects with limited budgets or focused research questions, WGBS and techniques like RRBS or enrichment-based methods could still be sufficient. But given our results, we now recommend the use of EM-Seq for better CpG coverage.

REFERENCES

1. Terzian et al. Benchmarking of sequencing technologies for CpG methylation detection: applications to quail and pig. Scientific data *in prep*
2. Simpson, J. T. et al. Detecting DNA cytosine methylation using nanopore sequencing. Nat. Methods 14, 407–410, 10.1038/nmeth.4184 (2017)
3. Liu, Y. et al. DNA methylation-calling tools for oxford nanopore sequencing: a survey and human epigenome-wide evaluation. Genome Biol.22, 10.1186/s13059-021-02510-z (2021)
4. Yuen, Z. W.-S. et al. Systematic benchmarking of tools for CpG methylation detection from nanopore sequencing. Nat. Commun.12, 10.1038/s41467-021-23778-6 (2021)
5. Sigurpalsdottir, B. D. et al. A comparison of methods for detecting dna methylation from long-read sequencing of human genomes. Genome Biol. 25, 10.1186/s13059-024-03207-9 (2024)
6. Ni et al. Benchmarking of Nanopore R10.4 and R9.4.1 flow cells in single-cell whole-genome amplification and whole-genome shotgun sequencing. Comput Struct Biotechnol J. 21:2352-2364. 10.1016/j.csbj.2023.03.038 (2023)
7. Eynard S. et al. Whole genome sequencing and epigenotyping for multiple generations of selection using Oxford Nanopore Technology. Environmental and Agronomical Genomics Symposium, Feb 2024, Toulouse, France. , pp.118, Book of abstracts. hal-04473354 (2024)
8. Cerutti C. et al. Transgenerational epigenetics in quail: whole genome DNA methylation analysis. Epigenetic Inheritance: Impact for Biology and Society, Aug 2023, Zurich, Switzerland. hal-04212971 (2023)

More about the SeqOccIn program : https://get.genotoul.fr/seqoccin/

Three majors topics are treated in this project: genome variability analysis, epigenetic mark analysis and metagenome analysis. It's carried out by Get-PlaGe and Genotoul Bioinfo platforms and financed by Feder funds (Programme Opérationnel FEDER-FSE_Midi-Pyrénées et Garonne 2014-2020). The project benefits from the contributions of INRAE research units GenPhySE, MIAT, GABI, GQE. 25 private partners are involved in the project.

Projet cofinancé par le Fonds Européen de Développement Régional