Alexei Zarubin[1], Elena Dementeva[2], Vladimir Elisashvili[3]

NanoMethCluster: A comprehensive tool for DNA methylation analysis using Oxford Nanopore sequencing data

*Affiliations
Institute of Microbial Biotechnology, Agricultural University of Georgia, Kakha Bendukidze Campus, 240 David Aghmashenebeli Alley, Tbilisi, 0159, Georgia
*Correspondence: Alexei Zarubin: a.a.zarubin@gmail.com

Abstract

Oxford Nanopore sequencing technology, capable of analyzing native DNA, has significantly advanced DNA methylation studies by enabling simultaneous nucleotide sequencing and methylation detection without bisulfite conversion. This advancement addresses the need to study cytosine methylation, a key epigenetic mechanism influencing gene regulation, development, and disease.

To fully utilize this technology, there is a growing demand for tools capable of performing a comprehensive analysis of methylation patterns directly from raw sequencing reads. Here, we present NanoMethCluster, a versatile software tool for extracting, normalizing, and clustering DNA methylation data, with advanced, user-friendly visualization capabilities.

The tool includes modules for read-level methylation data processing, dimensionality reduction, and clustering, enabling researchers to uncover epigenetic patterns across diverse biological contexts. The high-resolution capability of NanoMethCluster was demonstrated through the analysis of tumor and normal cell samples, as well as mixed-species data.

In this proof-of-concept study, these applications demonstrate their effectiveness in resolving complex methylation profiles and in providing useful information about biological variability and species-specific methylation patterns, while highlighting the need for validation in larger patient cohorts and additional species to fully assess their robustness and generalizability.

Keywords

DNA methylation, Oxford Nanopore sequencing, k-mer analysis, epigenetic clustering.

Introduction

The Oxford Nanopore sequencing technology has emerged as an accessible platform and reasonably accurate method (Q ≈ 20) for sequencing nucleic acids. This technology facilitates the sequencing of native DNA, enabling simultaneous analysis of nucleotide sequences and extraction of cytosine methylation data, including the identification of methyl- and hydroxymethyl-cytosines without bisulfite sequencing. Over the past decade,

Submitted/accepted for publication: December 5, 2025/January 29, 2026

sequencing technologies have complemented earlier single-gene efforts, ultimately providing a global understanding of DNA methylation and its dynamics in development and disease (Mattei et al., 2022). Cytosine methylation is a crucial epigenetic mechanism regulating gene expression, playing a vital role in various biological processes, including cellular differentiation, development, and responses to environmental stimuli. Notably, it allows for the reversion of somatic cells to their totipotent state (Li & Zhang, 2014). Although most endogenous promoters are not primarily regulated by methylation and methylation status can depend on cultivation conditions (Antequera et al., 1990; Borgel et al., 2010), methyl-cytosine (5mC) is often associated with gene expression repression, while hydroxymethyl-cytosine (5hmC) participates in active demethylation and can mark transcriptionally active regions of the genome (Stoyanova et al., 2021). DNA methylation plays a general regulatory role across different organisms. In plants, DNA methylation regulates development and environmental adaptation. In bacteria, DNA methylation is involved in gene expression regulation and replication, as the unmethylated strand of DNA cannot serve as a template in the subsequent round of replication (Gao et al., 2023).

In fungal pathogens, for instance, Fusarium species, which cause significant crop losses globally, epigenetic modifications regulate the expression of virulence genes (Ma et al., 2025). Soil microbiota epigenetic profiles directly influence plant health and productivity through complex interactions between microbes, indigenous microbiota, and plants (Chen et al., 2022).

Additionally, methylation serves as protection against phages via the restriction-modification system (Kang et al., 2025). The most common forms of methylation in bacteria are adenine methylation (N6-methyladenine) and cytosine methylation (5-methylcytosine) (Sánchez-Romero & Casadesús, 2020).

In viruses, methylation can affect their ability to infect host cells; for instance, hepatitis B virus (HBV) DNA methylation is associated with genomic expression regulation and potential malignancy (Koumbi & Karayiannis, 2016).

In fungi, DNA methylation is involved in regulating development, differentiation and responds to stress conditions. One proposed mechanism suggests that 5mC significantly affects DNA shape, impacting transcription factors' access to target sequences (Rao et al., 2018).

In mammals, methylation impacts numerous processes, including embryonic development, gene imprinting, and X-chromosome inactivation in females (Lee & Bartolomei, 2013; Alfeghaly & Rougeulle, 2025). Additionally, changes in methylation patterns are often observed in various diseases, particularly cancer. In cancer cells, hypermethylation of tumor suppressor gene promoters and hypomethylation of repeat-rich regions are common, contributing to genomic instability and oncogene activation (Pfeifer, 2018). Methylation analysis in the context of cancer helps identify epigenetic markers that can serve as diagnostic and prognostic tools and potential therapeutic targets. Despite advances in sequencing technologies, a growing number of programs now support methylation calling and downstream analysis, typically focusing either on extracting methylation probabilities from

raw signals or on aggregating site-level methylation rates at the genome level. However, most tools focus either on site-level calling from raw signals (e.g., Dorado modkit) or on aggregating methylation rates at the genome level. Few tools provide an end-to-end framework for working directly with raw reads and their k-mer context, which is essential for capturing read-to-read heterogeneity and subtle but biologically significant epigenetic differences.

To address this gap, we present NanoMethCluster, a new tool developed for analyzing DNA methylation at the level of raw reads. The software includes modular components for extracting methylation calls from BAM files, filtering and normalizing read- and k-mer-level features, performing dimensionality reduction, clustering reads, and visualizing the results in an interactive format. NanoMethCluster is designed to enhance the interpretability and efficiency of methylation analysis, facilitating deeper insights into epigenetic regulation across diverse biological contexts. To provide a preliminary evaluation of NanoMethCluster, we applied its core modules to two representative use cases: a tumor versus normal human cell line dataset and a mixed species metagenomic sample.

Materials and Methods

Sample Preparation and DNA Extraction

For testing and demonstration, a mixed sample of Ewingella americana and *Fusarium fujikuroi* was used in a 50:50 ratio obtained in our laboratory, as well as publicly available Nanopore sequencing data from normal (COLO829BL) and tumor (COLO829) human cell lines. Frozen mycelium and bacterial cultures were ground into fine powder using a mortar and pestle in liquid nitrogen.

DNA was extracted based on a modified free-column extraction method suitable for long-read sequencing (Jones et al., 2020) with incubation in lysis buffer and RNAse A treatment.

The PVP-based size selection method was used to increase the recovery of high-molecular-weight fragments based on the protocol approved for Nanopore sequencing (Oxford Nanopore Technologies, 2024). Sample integrity and quantity were assessed on an agarose gel (Bento Lab, Bento Bioworks Ltd.) and fluorometrically using the QuDye dsDNA BR Assay Kit (Lumiprobe).

Nanopore Sequencing and Read Processing

The DNA library was prepared from enriched DNA (*Ewingella americana* and *Fusarium fujikuroi* laboratory strains), barcoded, and pooled using the manufacturer's protocols for the Native Barcoding Kit 24 (SQK-NBD114, Oxford Nanopore Technologies).

The library was sequenced on a MinION flow cell (R10.4.1, Oxford Nanopore Technologies) using MinKNOW Sequencing software (version 23.11.5, Oxford Nanopore Technologies).

A rapid assessment of species diversity was performed using Kraken2 (Wood et al., 2019) integrated into EPI2ME with the database k2_pluspf_16gb_20240112. Basecalling and adapter trimming were conducted using Guppy (version 3.2.2) after sequencing (Oxford Nanopore Technologies, 2024).

NanoMethCluster Modules

NanoMethCluster
https://github.com/alekseizarubin/NanoMethCluster
comprises four software modules:

BAMReadAnalyzer

BAMReadAnalyzer extracts methylation information from BAM files obtained using the Dorado program and the mC_5hmC and 5mCG_5hmCG models. It provides data on read length, nucleotide composition, and the number of methylated positions in each read. Users can analyze total mC+hmC methylation or each type of methylation separately.

1. Base Method: This method assigns a specific modification type to a position based on the highest probability, according to the ML flag in the BAM file. It yields more contrasting values.
2. Probability Model: This method uses the Monte Carlo method to account for the contribution of even small probabilities to potential methylation levels. It can be controlled by specifying a seed number for reproducibility and the number of simulation cycles, with recommended values between 100 and 10,000 cycles. More cycles increase accuracy, especially for long reads, but require significant computational resources. Multithreading is implemented to address execution time issues. The basic output can be used to assess sample methylation levels. Additionally, a k-mer counting method is implemented to analyze methylation in the context of nucleotide sequences. K-mer frequency analysis is essential for studying DNA methylation, as k-mers represent short sequences of k nucleotides that can identify unique and repetitive elements in the genome or metagenome. This information is useful for subsequent read clustering within a single organism or in metagenomic studies. The frequency of methylated k-mers helps assess methylation patterns in various genomic contexts.

KmerDataProcessor: Normalization and Filtering of Reads

The KmerDataProcessor module is designed for filtering reads based on their average quality and length, as well as for normalizing k-mers. This process is critical for preparing data before further DNA methylation analysis. The module implements three options for k-mer normalization: merging complementary k-mers, merging direct and reverse read variants, and combining both complementary and direct/reverse reads.

1. Filtering Reads: Reads are filtered based on minimum length and average quality thresholds. Reads that do not meet the specified criteria are excluded from further analysis.
2. Normalization of Nucleotides and Counters: After filtering, nucleotide data are normalized by the length of the reads. This means that the count of each nucleotide (A, T, G, C) is divided by the read length to obtain relative frequencies. Similarly, methylation counters (e.g., Count_Max_Ch, Count_Max_Cm) are normalized by the number of base pairs (Num_Pairs).
3. K-mer Normalization: The module identifies columns containing k-mer data and normalizes them by the length of the reads. The three normalization options are:

a.Merging complementary k-mers (Complementary, 'c'): Each k-mer is merged with its complementary k-mer.

These values are summed, and one of the columns is removed.

b. Merging direct and reverse read variants (Reverse, 'r'): K-mers are merged with their reverse sequences.

c. Merging both complementary and reverse

variants (Complementary and Reverse, 'cr'): K-mers are merged with both their complementary and reverse sequences.

Data Analyzer: Dimensionality Reduction and Clustering

The DataAnalyzer module is designed for dimensionality reduction using PCA (Principal Component Analysis) and clustering using K-means. It operates on the normalized feature tables produced by KmerDataProcessor, including read-level methylation counters and k-mer frequencies. This module evaluates the optimal number of components using the elbow method and provides tables with results from t-SNE (t-Distributed Stochastic Neighbor Embedding) and UMAP (Uniform Manifold Approximation and Projection) analyses.

1. PCA Analysis: PCA is used to reduce the dimensionality of the data while retaining the most important variance aspects. This is achieved by transforming the original features into a smaller number of principal components, which explain the maximum amount of variance in the data. The module offers an "auto" option to determine the number of components or allows the user to specify the number.

2. t-SNE Analysis: t-SNE is used for visualizing high-dimensional data by reducing it to two or three dimensions. This technique is particularly effective for embedding high-dimensional data into a low-dimensional space of two or three dimensions, which can then be visualized in scatter plots.

3.UMAP Analysis: UMAP is another dimensionality reduction technique that is used for embedding high-dimensional data into low-dimensional space, preserving more of the global structure compared to t-SNE.

4. K-means Clustering: K-means clustering is a method used to partition the data into K distinct clusters. This module determines the optimal number of clusters by evaluating the distortions for different values of K and using the elbow method to find the optimal number of clusters.

PlotGenerator: Interactive Visualization and Summary Tables

The PlotGenerator module is designed to generate interactive plots using Plotly for the visual analysis of results and to create summary tables that assess the average methylation levels in individual clusters. This module also integrates analysis results with Kraken2 outputs for a comprehensive metagenomic sample analysis.

1. Interactive Plot Generation: This includes reading gzipped TSV files containing the PCA, t-SNE, and UMAP results, as well as Kraken2 files for taxonomic classification. The module creates interactive 2D or 3D scatter plots from the PCA, t-SNE, and UMAP results, with optional cluster coloring.

2. Summary Table Generation: The module reads the K-means clustering results, calculates the average methylation levels within each cluster, and generates summary tables and matrices showing the distribution of reads across different clusters and Kraken labels.

Results

Using NanoMethCluster, we analyzed two datasets: a cancer dataset comprising normal (COLO829BL) and tumor (COLO829) samples, and a custom dataset containing a mix of Fusarium fujikuroi and Ewingella americana. The clustering results were compared to those obtained using Kraken, highlighting NanoMethCluster's capability for resolving

methylation patterns. These two datasets should be regarded as proof-of-concept use cases, and systematic benchmarking on additional cancer types, tissues, and metagenomic communities will be required to generalize these observations.

Normal and Tumor Samples

NanoMethCluster can be used both for analyzing samples from different species and for clustering samples obtained from the same individual that differ in their methylation profiles, for example, tumor versus normal samples. We used 1% random long reads from COLO829 and COLO829BL samples (Oxford Nanopore Technologies, 2023). Instead of using Kraken2, we loaded a similarly structured file where the sample to which each read belongs was specified instead of taxon data. Using the **BAMReadAnalyzer** module, methylation and hydroxymethylation data were extracted to enable a detailed comparison of methylation profiles between tumor and normal samples. Interactive visualizations created by the **PlotGenerator** module further enhanced understanding by displaying methylation distributions within the clusters.

The specific commands used are provided in the supplementary materials. Clear clustering according to methylation levels is observed, with normal sample readings showing more contrasting methylation levels (Table 1, Fig. 1). Specifically, 45% of all reads from the normal sample are in the most contrasting clusters (2 and 3), while only 12% of the reads from the tumor sample fall into these clusters. The tumor sample is characterized by less contrasting and more demethylated average methylation levels, with 45% of reads falling into clusters 0 and 8, with methylation levels of 0.26 and 0.4, respectively.

Mixed Sample of Two Species

In the mixed-species dataset containing Fusarium fujikuroi and Ewingella americana, NanoMethCluster effectively resolved species-specific methylation patterns. Using the **BAMReadAnalyzer** module, we identified differences in methylation and hydroxymethylation levels between the two species. Fusarium fujikuroi exhibited higher hydroxymethylation in specific clusters, whereas Ewingella americana displayed more uniform methylation levels (Table 2, Fig. 2). The **KmerDataProcessor** module normalizes k-mer counts to ensure balanced data representation across species, facilitating reliable clustering. Dimensionality reduction and clustering with the **DataAnalyzer** module revealed distinct groups corresponding to each species, and interactive visualizations generated by the **PlotGenerator** module highlighted methylation gradients within clusters, further supporting species-specific differences.

Clear clustering based on species is observed, with a gradient of average methylation levels within each major cluster.

Discussion

NanoMethCluster provides a robust and user-friendly platform for analyzing DNA methylation directly from raw Nanopore sequencing reads and focus on read-level clustering combined with k-mer contextualization.

By integrating modules for filtering, normalization, clustering, and visualization, it offers comprehensive insights into methylation dynamics across different biological contexts.

Our proof-of-concept applications to cancer and mixed-species data demonstrate that read-level methylation features, combined with dimensionality reduction and clustering, can reveal biologically meaningful structure that is not readily apparent from site-level summaries alone.

In the normal and tumor samples (Table 1), NanoMethCluster revealed that 45% of normal sample reads clustered into the most contrasting methylation groups (clusters 2 and 3, with 5mC ~0.73–0.90), while 45% of tumor reads fell into clusters with lower average methylation (clusters 0 and 8, with 5mC ~0.26–0.40). This demonstrates read-level separation based on methylation profiles.

These findings are consistent with the notion that cancer methylomes often exhibit global hypomethylation and altered distribution of methylated cytosines, and they illustrate how read-level clustering can help summarize such shifts at the sample level.

In the mixed-species dataset (Table 2), NanoMethCluster distinguished between Fusarium fujikuroi and Ewingella americana with high accuracy (for example Cluster 8: 95.6% Ewingella), highlighting its suitability for analyzing complex metagenomic samples. Species-enriched clusters were characterized by distinct 5mC and 5hmC profiles, suggesting that methylation-aware clustering may provide an additional layer of information for resolving community structure and for prioritizing taxa or genomic regions with unusual methylation patterns.

A comparison with Kraken2 highlights how NanoMethCluster complements taxonomic classification by adding an orthogonal layer of information based on methylation patterns rather than sequence composition alone.

For example, in the analysis of normal and tumor samples, NanoMethCluster revealed that 45% of normal sample reads clustered into the most contrasting methylation groups (clusters 2 and 3), while tumor reads were enriched in clusters with lower average methylation levels (clusters 0 and 8). Similarly, in the mixed-species dataset, NanoMethCluster produced distinct, species-enriched clusters for Fusarium fujikuroi and Ewingella americana that were consistent with Kraken2 assignments and further characterized by their methylation profiles. Rather than replacing Kraken2, NanoMethCluster integrates read-level methylation profiling with taxonomic information to facilitate more detailed exploration of epigenetic heterogeneity in complex samples.

At the same time, the present work should be regarded as a preliminary evaluation of NanoMethCluster. The tumor-normal use case is limited to a single cell line pair, and the metagenomic example comprises only one two-species mixture.

Systematic benchmarking across larger and more diverse patient cohorts, additional cancer types, and more complex microbial communities will be needed to assess the robustness and generalizability of the observed clustering patterns. Future comparisons with alternative methylation analysis frameworks will also be important to better delineate the specific advantages and limitations of the read-level clustering approach implemented in NanoMethCluster.

Future developments of NanoMethCluster aim to improve its computational efficiency, extend systematic validation across broader panels of species and clinical cohorts, and introduce new

features, such as advanced taxonomic profiling and enhanced predictive tools for methylation patterns. These developments are expected to make NanoMethCluster even more versatile and reliable for addressing complex epigenetic research challenges and for integrating methylation-aware clustering into standard analysis workflows for Nanopore data.

Conclusion

In summary, NanoMethCluster provides an end-to-end, user-friendly framework for analyzing DNA methylation directly from raw Oxford Nanopore reads, from feature extraction and normalization to clustering and interactive visualization.

In proof-of-concept applications, the tool successfully separated normal and tumor samples based on differences in methylation profiles and resolved species-specific methylation patterns in a mixed Fusarium fujikuroi-Ewingella americana dataset, complementing Kraken2-based taxonomic assignments.

Although these results are based on a single tumor-normal pair and one two-species mixture and should therefore be considered preliminary, they illustrate the potential of read-level methylation clustering to provide an orthogonal view of epigenetic heterogeneity in both clinical and metagenomic contexts. With further optimization and validation across larger and more diverse datasets, NanoMethCluster may become a useful component of standard pipelines for Nanopore-based DNA methylation analysis.

Acknowledgment: The research was supported by the ISTC Project No. PR 132.1.1

Conflicts of Interest: The authors declare that they have no conflicts of interest.

References

Mattei, A. L., Bailly, N., & Meissner, A. (2022). DNA methylation: A historical perspective. Trends in Genetics, 38(7), 676–707. https://doi.org/10.1016/j.tig.2022.03.010

Li, E., & Zhang, Y. (2014). DNA methylation in mammals. Cold Spring Harbor Perspectives in Biology, 6(5), a019133. https://doi.org/10.1101/cshperspect.a019133

Antequera, F., Boyes, J., & Bird, A. (1990). High levels of De Novo methylation and altered chromatin structure at CpG islands in cell lines. Cell, 62(3), 503–514. https://doi.org/10.1016/0092-8674(90)90015-7

Borgel, J., Guibert, S., Li, Y., Chiba, H., Schübeler, D., Sasaki, H., Forné, T., & Weber, M. (2010). Targets and dynamics of promoter DNA methylation during early mouse development. Nature Genetics, 42(12), 1093–1100. https://doi.org/10.1038/ng.708

Stoyanova, E., Riad, M., Rao, A., & Heintz, N. (2021). 5-Hydroxymethylcytosine-mediated active demethylation is required for mammalian neuronal differentiation and function. eLife, 10, e66973. https://doi.org/10.7554/eLife.66973

Gao, Q., Lu, S., Wang, Y., He, L., Wang, M., Jia, R., Chen, S., Zhu, D., Liu, M., Zhao, X., Yang, Q., Wu, Y., Zhang, S., Huang, J., Mao, S., Ou, X., Sun, D., Tian, B., & Cheng, A. (2023). Bacterial DNA methyltransferase: A key to the epigenetic world with lessons learned from proteobacteria. Frontiers in Microbiology, 14. https://doi.org/10.3389/fmicb.2023.1129437

Ma, Y., Wang, X., & Li, X. (2025). The emerging role of DNA methylation in the pathogenicity of bacterial pathogens. Journal of Bacteriology, 207(8), e00108-25. https://doi.org/10.1128/jb.00108-25

Chen, C., Wang, M., Zhu, J., Tang, Y., Zhang, H., Zhao, Q., Jing, M., Chen, Y., Xu, X., Jiang, J., & Shen, Z. (2022). Long-term effect of epigenetic modification in plant-microbe interactions: Modification of DNA methylation induced by plant growth-promoting bacteria mediates promotion process. Microbiome, 10, 36. https://doi.org/10.1186/s40168-022-01236-9

Kang, H., Gao, A., & Zhu, Y. (2025). Bacterial restriction-modification systems: Mechanisms of defense against phage infection. Biophysics Reports, 11(5), 330–343. https://doi.org/10.52601/bpr.2025.240070

Sánchez-Romero, M. A., & Casadesús, J. (2020). The bacterial epigenome. Nature Reviews. Microbiology, 18(1), 7–20. https://doi.org/10.1038/s41579-019-0286-2

Koumbi, L., & Karayiannis, P. (2016). The Epigenetic Control of Hepatitis B Virus Modulates the Outcome of Infection. Frontiers in Microbiology, 6. https://doi.org/10.3389/fmicb.2015.01491

Rao, S., Chiu, T.-P., Kribelbauer, J. F., Mann, R. S., Bussemaker, H. J., & Rohs, R. (2018). Systematic prediction of DNA shape changes due to CpG methylation explains epigenetic effects on protein-DNA binding. Epigenetics & Chromatin, 11(1), 6. https://doi.org/10.1186/s13072-018-0174-4

Lee, J. T., & Bartolomei, M. S. (2013). X-Inactivation, Imprinting, and Long Noncoding RNAs in Health and Disease. Cell, 152(6), 1308–1323. https://doi.org/10.1016/j.cell.2013.02.016

Alfeghaly, C., & Rougeulle, C. (2025). X chromosome inactivation in mammals: General principles and species-specific considerations. EMBO reports, 26(14), 3478–3490. https://doi.org/10.1038/s44319-025-00499-1

Pfeifer, G. P. (2018). Defining Driver DNA Methylation Changes in Human Cancer. International Journal of Molecular Sciences, 19(4), 1166. https://doi.org/10.3390/ijms19041166

Jones, A., Torkel, C., Stanley, D., Nasim, J., Borevitz, J., & Schwessinger, B. (2020). Scalable high-molecular weight DNA extraction for long-read sequencing. https://www.protocols.io/view/scalable-high-Molecular-weight-dna-extraction-for-bnjhmcj6

Oxford Nanopore Technologies. (2024a). Size selection of HMW DNA by semi-selective DNA precipitation. http://community.nanoporetech.com/extraction_methods/size-selection2

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. Genome Biology, 20(1), 257. https://doi.org/10.1186/s13059-019-1891-0

Oxford Nanopore Technologies. (2024b). Guppy basecaller. https://nanoporetech.com/community/downloads

Oxford Nanopore Technologies. (2023). ONT Data colo829_2023.04. s3://ont-open-data/colo829_2023.04.

Table 1. Comparison of methylation levels between tumor and normal samples across clusters. For each cluster, the table reports average 5hmC and 5mC fractions calculated from the readings assigned to that cluster in the tumor and normal samples.

Cluster	Methylation level		Average	
	Tumor samples	**Normal samples**	**5hmC**	**5mC**
0	0.206464	0.068246	0.021683	0.255298
1	0.088298	0.171441	0.039625	0.728413
2	0.097962	0.253046	0.023445	0.89515
3	0.026871	0.205767	0.01434	0.055871
4	0.086711	0.034181	0.039773	0.443376
5	0.175377	0.119647	0.038939	0.566413
6	0.011085	0.00758	0.063659	0.478603
7	0.010398	0.006447	0.061333	0.454381
8	0.252034	0.057486	0.025694	0.399628
9	0.044801	0.076159	0.036895	0.76532

5hmC: Average fraction of hydroxymethylated cytosines detected in the reads within a particular cluster. 5mC: Average fraction of methylated cytosines detected in the reads within a particular cluster.

Table 2. Methylation levels in mixed-species samples. The table shows the distribution of reads across clusters and the corresponding average 5hmC and 5mC fractions for Fusarium fujikuroi and Ewingella americana.

	Read Count		**Cluster-Level Methylation**	
Cluster	***Fusarium fujikuroi***	***Ewingella americana***	**5hmC**	**5mC**
0	9	8	0.070789	0.709826
1	10971	3595	0.038531	0.057623
2	10	2412	0.078449	0.109559
3	7282	2762	0.0358	0.050849
4	11651	4752	0.05158	0.067641
5	130	393	0.027893	0.036271
6	487	10617	0.065496	0.088448
7	413	0	0.012206	0.02194
8	481	10565	0.050348	0.076395

Read counts are Kraken2-based per-species counts per cluster; 5mC/5hmC are cluster-level means across all reads in the cluster.

5hmC: Average fraction of hydroxymethylated cytosines detected in the reads within a particular cluster.

5mC: Average fraction of methylated cytosines detected in the reads within a particular cluster.

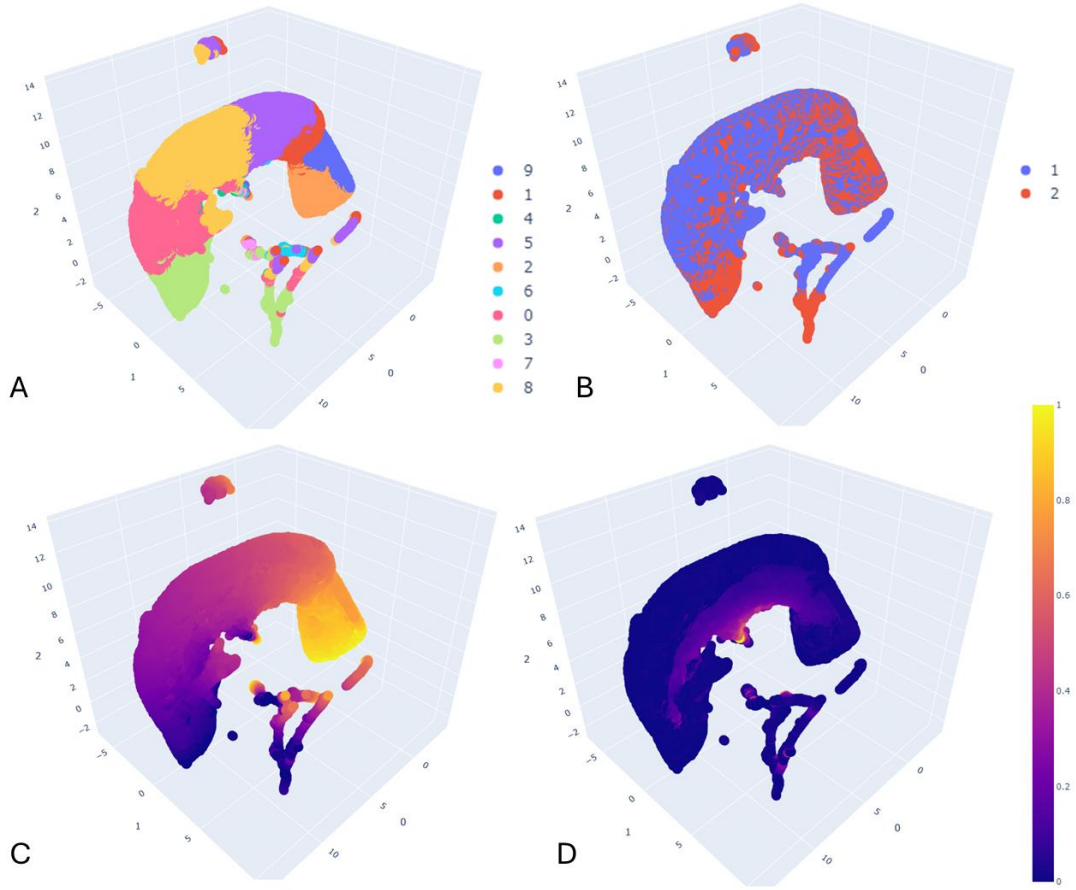

Fig. 1. Clustering of normal and tumor samples based on methylation levels.

A: Results of K-means clustering of long-read methylation data (clusters shown in distinct colors).

B: Sample origin distribution (1: tumor samples; 2: normal samples, color-coded).

C: Distribution of methylation levels across clusters (color gradient: 0 - unmethylated, 1 - fully methylated).

D: Distribution of methylation levels across clusters (color gradient: 0 - no hydroxymethylation, 1 - fully hydroxymethylated).

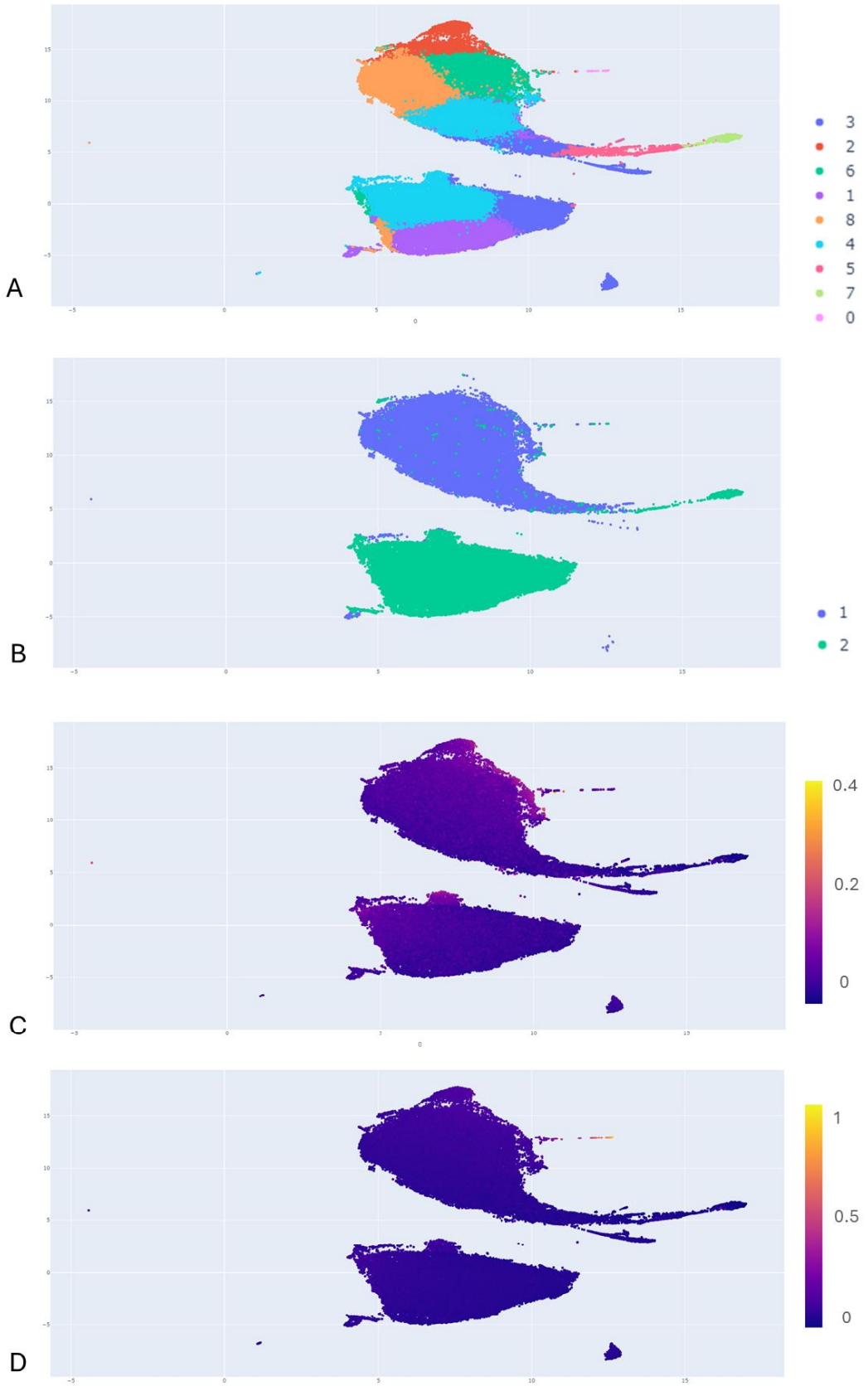

Fig. 2. Clustering of mixed-species samples based on methylation levels.
A: Results of K-means clustering of methylation data (clusters shown in distinct colors).

B: Species origin classification (color-coded: 1 - Ewingella americana and 2 - Fusarium fujikuroi, based on Kraken2 results).
C: Distribution of methylation levels per cluster (color gradient: 0 to 0.4).
D: Distribution of methylation levels per cluster (color gradient: 0 to 1).
The results provide initial evidence of NanoMethCluster's ability to resolve complex methylation patterns and support its potential application across diverse biological contexts.