Data in Brief 65 (2026) 112522

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Metabarcoding data: Full-length 16S rRNA sequence of endophytic bacteria in the root of asymptomatic and blast-symptomatic rice plants (*Oryza sativa*, L.)

Yasir Sidiq[a,d,*], Triastuti Rahayu[a,d], Peni Indrayudha[b,d], Erma Musbita Tyastuti[a,d], Azmi Zaki Waliudin Althaf[c], Banuwati Kartika Sari[a]

[a] *Department of Biology Education, Universitas Muhammadiyah Surakarta, Surakarta, Indonesia*
[b] *Pharmacy Department, Universitas Muhammadiyah Surakarta, Surakarta, Indonesia*
[c] *Department of Biotechnology, Graduate School of Universitas Gadjah Mada, Yogyakarta, Indonesia*
[d] *Microbial Engineering Research Center, Universitas Muhammadiyah Surakarta, Surakarta, Indonesia*

ARTICLE INFO

Article history:
Received 11 November 2025
Revised 17 January 2026
Accepted 21 January 2026
Available online 27 January 2026

Dataset link: PRJNA992961 (Original data)

Keywords:
Barcoding gene
Endophytic bacteria
Full-length sequence
Rice root

ABSTRACT

There is a sustained demand for biofertilizers to enhance crop productivity. Endophytic bacteria associated with disease-tolerant rice varieties offer significant potential as biofertilizers; however, the bacteriome diversity within these plants remains underexplored. This dataset presents full-length 16S metagenomic sequences of endophytic bacteria isolated from the roots of blast-infected and uninfected rice plants. Root samples were processed and subjected to surface sterilisation. Following total genomic DNA extraction, sequencing was performed using 16S ribosomal RNA primers via the high-throughput Oxford Nanopore Technologies platform. The raw sequence data were filtered for quality control using NanoFilt. Subsequently, the sequences were aligned against the National Center for Biotechnology Information (NCBI) 16S RefSeq database to identify the species of the endophytic root bacteria. The data associated with this project have been registered in the NCBI BioProject database under accession number PRJNA992961. The dataset comprises two

* Corresponding author.
E-mail address: ys120@ums.ac.id (Y. Sidiq).

https://doi.org/10.1016/j.dib.2026.112522

distinct sample groups, each analysed in duplicate, with sequencing yields ranging from 17.7 to 20.3 Mb. Consequently, this dataset provides valuable insights regarding the comparative composition of endophytic bacteria inhabiting healthy roots versus those found in blast-infected rice. Characterizing this diversity, particularly within healthy rice plants, is essential for foundational research underpinning the future development of biofertilizers.

Specifications Table

Subject	Biology
Specific subject area	Root endophytic bacterial communities associated with rice (*Oryza sativa* L.)
Type of data	Raw FASTQ sequence files, quality-controlled sequence data, taxonomic assignment tables, metadata files. And figures (bar charts, PCoA plots, Krona plots).
Data collection	Surface-sterilized rice root tissues were subjected to genomic DNA extraction, followed by PCR amplification of the full-length bacterial 16S rRNA gene. Amplicon were sequenced using Oxford Nanopore Technologies to generate long-read 16S rRNA sequence data in FASTQ format.
Data source location	Surakarta City, Central Java, Indonesia, (7°32′18″S 110°47′29″E)
Data accessibility	Repository name: NCBI Sequence Read Archive (SRA) Data identification number: PRJNA992961 Direct URL to data: https://www.ncbi.nlm.nih.gov/sra/PRJNA992961 Instructions for accessing these data: https://www.ncbi.nlm.nih.gov/sra/docs/sradownload/
Related research article	none

1. Value of the Data

- This study characterizes the comprehensive full-length 16S rRNA gene metasequences of bacterial assemblages inhabiting the roots of asymptomatic and blast-symptomatic rice plants via high-fidelity long-read sequencing.
- The dataset affords high-resolution taxonomic classification, a level of granularity seldom achievable in extant rice blast literature and consequently establishes a unique reference repository for the discipline.
- The data reveal marked disparities in the diversity indices and community structure of endophytic populations distinguishing healthy from blast-infected roots.
- This repository possesses significant utility for the wider scientific community by enabling robust secondary meta-analyses and facilitating comparative investigations into endophytic bacteriome dynamics.
- These sequences enable the targeted bioprospecting of putative beneficial taxa exhibiting antagonistic potential against blast pathology to aid the discovery of novel biocontrol agents.
- The dataset underpins translational applications ranging from the identification of bacterial diagnostic biomarkers and the design of specific primers to the formulation of synthetic bacterial consortia for enhanced crop protection strategies.

2. Background

The investigation of beneficial endophytic bacteria capable of enhancing crop growth and yield constitutes a pivotal area of research. Residing within internal plant tissues, these

Table 1
Summary of sequencing quality metrics and taxonomic classification success for asymptomatic and blast-symptomatic rice root samples.

Samples	Filtered Reads	Mean Read Length (bases)	Mean Q Score	>Q10 (%)	>Q12 (%)	>Q15 (%)	Classification Success (%)
Asymptomatic_1	12,395	1548.7	12	87.9	51.2	1.6	99.39
Asymptomatic_2	12,370	1546.2	12	87.8	51.1	1.7	99.45
Blast-symptomatic_1	13,046	1553.3	12	87.3	51.7	1.8	99.49
Blast-symptomatic_2	11,478	1545.1	12	88.0	51.1	1.7	99.09

microorganisms facilitate growth through diverse mechanisms, including phytohormone synthesis, biological nitrogen fixation, phosphorus solubilisation, and the modulation of plant defence responses [1]. In a previous study, we characterized the core endophytic bacterial communities associated with disease-tolerant banana plants, differentiating them from populations found in susceptible cultivars [2]. However, such investigations are frequently constrained by the limitations of culture-dependent methods, as the vast majority of bacterial taxa are recalcitrant to cultivation on conventional laboratory media [3]. Consequently, metabarcoding provides a robust, culture-independent methodology for the comprehensive characterization of the total microbiome inhabiting specific plant tissues.

Comparative analyses of endophytic communities inhabiting tolerant and susceptible plants facilitate the identification of key bacterial taxa associated with disease resistance. A substantial body of research, including studies conducted in Vietnam, has demonstrated significant compositional divergence contingent upon host genotype and environmental conditions [4–7]. Nevertheless, investigations of this nature remain sparse within the Indonesian context. Consequently, the present dataset offers a comparative analysis of the endophytic bacterial communities colonising the roots of asymptomatic and blast-symptomatic rice (*Oryza sativa* L.) plants.

3. Data Description

This dataset comprises full-length 16S rRNA gene sequencing data generated from endophytic bacterial communities colonising the roots of asymptomatic and blast-symptomatic rice plants. Root samples were collected from field-cultivated rice, encompassing both asymptomatic and blast-symptomatic hosts. Sequencing was performed using Oxford Nanopore Technologies (ONT), producing long read 16S rRNA amplicon data.

As detailed in Table 1, the assessment of sequence data quality indicates that all samples yielded filtered full-length 16S rRNA reads ranging from 11,478 to 13,046, with a mean read length of approximately 1.55 kb. The mean Phred quality (Q) score across samples was 12. Regarding quality thresholds, 87–88 % of reads met the Q10 standard, approximately 51 % achieved Q12, and 1.6–1.8 % exceeded Q15. Moreover, the taxonomic classification success rate was 99.1–99.5 %.

Chao1 richness values for asymptomatic and blast-symptomatic samples are shown in Fig. 1A. The first three principal axes described 62.7 % (PC1), 22.6 % (PC2), and 14.8 % (PC3) of the variance (Fig. 1B).

Fig. 2 describes species level abundance profiles derived from full-length 16S rRNA sequencing of the roots of asymptomatic and blast-symptomatic rice plants. A total of ten species were identified across both sample groups. The figure displays the distribution of 11 taxonomic categories**,** comprising 10 named bacterial taxa (*Limnospira fusiformis, Sphingomonas panni, Euryhalinema mangrovii, Hydrogenispora ethanolica, Olisthonema eestii, Dulcicalothrix alborzica, Sphingomonas carotinifaciens, Komarekiella gloecapsioidea, Acinetobacter albensis*, and *Aegeococcus thureti*) and one aggregated category ("Others")**.** Actual abundance values are shown for each biological replicate within each sample group.

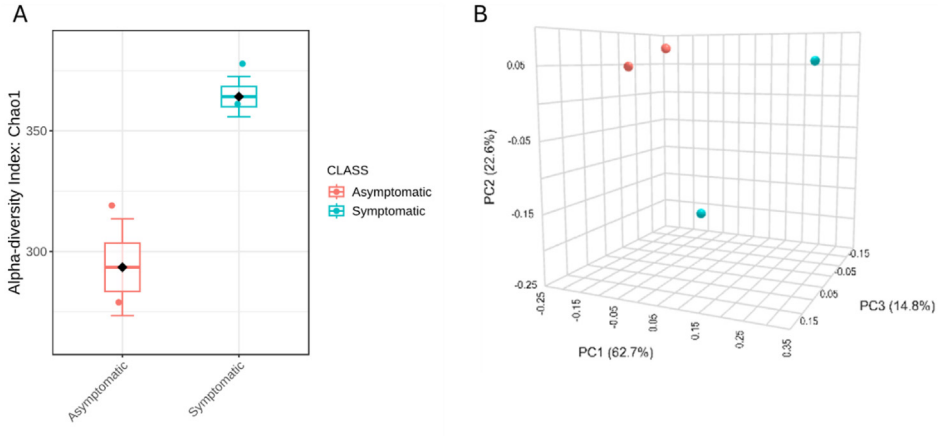

Fig. 1. Diversity and community structure of the rice root endophytic bacteriome. (A) Boxplot illustrating alpha-diversity (Chao1 richness index) in asymptomatic (red) and blast-symptomatic (cyan) root samples. Individual data points represent biological replicates, while diamonds denote group means. (B) Three-dimensional Principal Coordinate Analysis (PCoA) plot showing the ordination of the root of asymptomatic (red) and blast-symptomatic (cyan) bacterial communities.

The Krona plots in Fig. 3 present the hierarchical taxonomic composition of endophytic bacterial communities. This analysis spans from the phylum to the species level in both asymptomatic and symptomatic rice roots. Separate Krona plots are shown for asymptomatic and symptomatic rice roots to represent the taxonomic structure of each sample group. Within each Krona plot, circular segments correspond to successive taxonomic ranks, including phylum, class, order, family, genus, and species, with segment sizes reflecting the abundance values associated with each taxon.

4. Experimental Design, Materials and Methods

4.1. Sample collection, DNA extraction, and 16S rRNA gene amplification

Rice root specimens were obtained from cultivation fields located in Surakarta, Central Java, Indonesia. To ensure representative sampling, three distinct root systems were pooled to constitute a single composite sample. This procedure was executed in duplicate for both symptomatic and asymptomatic experimental groups (Fig. 4). Prior to DNA extraction, root tissues underwent a rigorous surface sterilisation protocol comprising immersion in 70 % (*v/v*) ethanol for 1 min followed by exposure to 2 % (*v/v*) sodium hypochlorite for 3 min under constant agitation. The process was completed with five successive rinses utilising sterile distilled water. To verify the efficacy of surface sterilisation, the final rinse water was plated onto nutrient agar [2]. Genomic DNA extraction was subsequently performed utilising the ZymoBIOMICS DNA Miniprep Kit (Zymo Research, D4300). Briefly, 100 mg of root tissue was transferred into a lysis tube containing ceramic beads and 750 µL of lysis buffer prior to homogenisation for 5 to 10 min via a bead beater. The resulting lysates were centrifuged at 10,000 × *g* for 1 min, and the supernatant was loaded onto a silica spin column. The DNA was bound to the column, washed twice with DNA Wash Buffer, and eluted in 50 µL of DNA Elution Buffer [8]. DNA concentration was quantified utilising a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific, USA). Library preparation was conducted utilising kits provided by Oxford Nanopore Technologies. Amplification of the full length 16S rRNA gene was performed using Phusion High Fidelity PCR Master Mix (New England Biolabs) [2]. Specifically, each 25 µL reaction mixture comprised

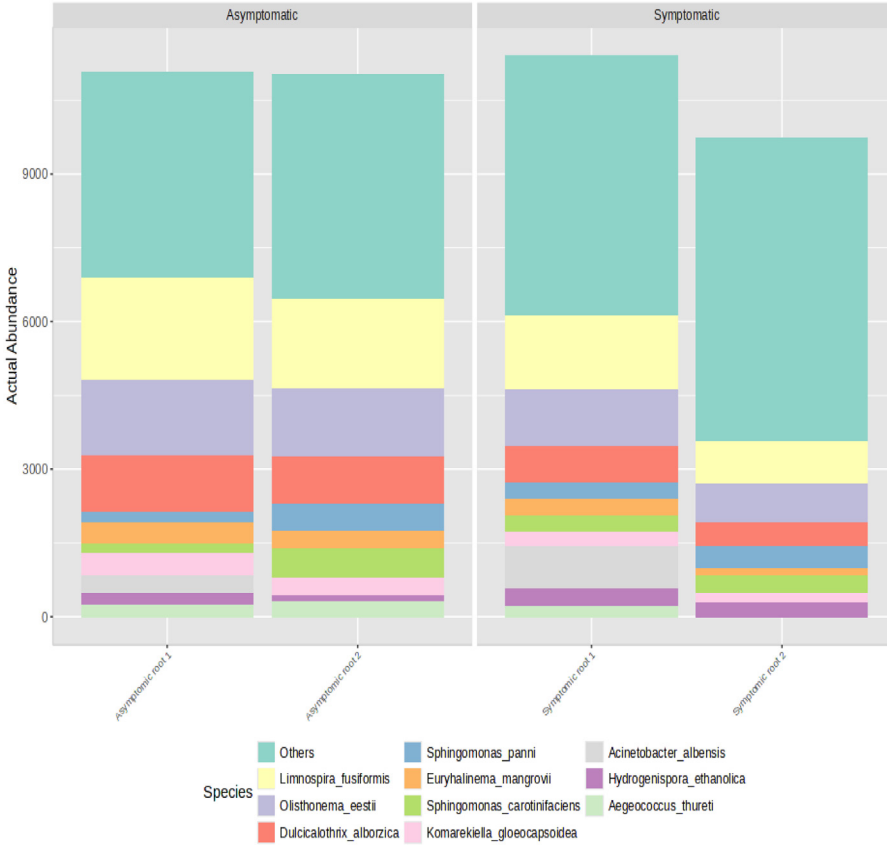

Fig. 2. Stacked bar plots illustrating the absolute abundance of dominant endophytic bacterial species within the roots of asymptomatic and blast-symptomatic rice plants.

12.5 µL of 2× Phusion Master Mix, 1 µL of both forward and reverse primers (10 µM), 2 µL of DNA template (5 to 10 ng), and water free of nucleases to adjust the final volume. The reaction utilised the primer pair of 27F-5′-AGA GTT TGA TCM TGG CTG AG-3′ and 1492R-5′-CGG TTA CCT TGT TAC GAC TT-3′[9]. The PCR reaction commenced with an initial denaturation at 98 °C for a duration of 2 min. Subsequently, the protocol proceeded through 35 cycles comprising an annealing stage of 15 s, wherein the temperature decreased gradually from 65 °C to 55 °C, followed by an extension phase of 30 s at 68 °C. The annealing temperature was reduced by 1 °C per cycle until reaching 55 °C [8].

4.2. 16S rRNA gene sequencing

Nanopore sequencing was conducted using MinKNOW software (version 22.05.7). Subsequent basecalling was performed via Guppy (version 6.1.5) using the high-accuracy model [10]. The quality of the resulting FASTQ files was assessed using NanoPlot, and quality filtering was executed using NanoFilt. A summary of the sequence quality control metrics was compiled into a table [11,12]. To monitor for potential contamination, a blank DNA extraction sample and a no-template PCR control were processed and sequenced alongside the experimental samples.

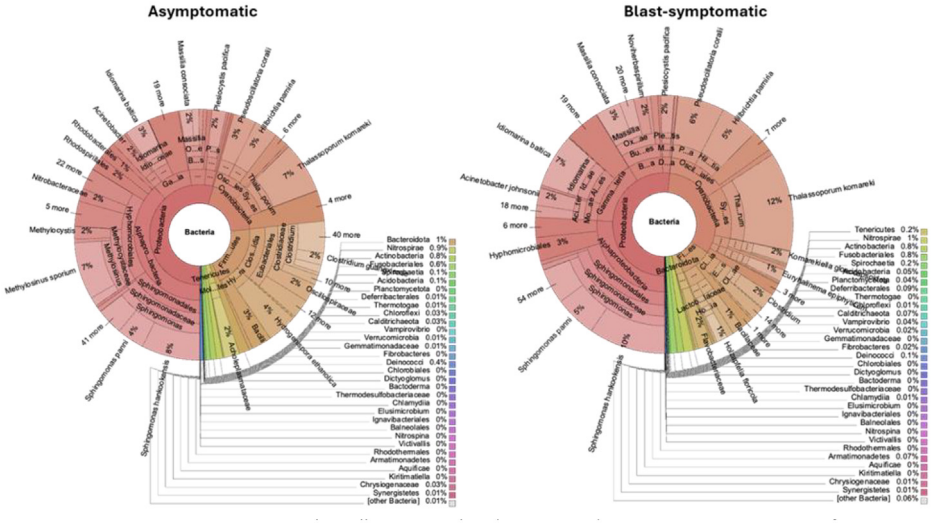

Fig. 3. Representative Krona plots illustrating the diversity and taxonomic composition of endophytic bacteria within the roots of asymptomatic and blast-symptomatic rice plants.

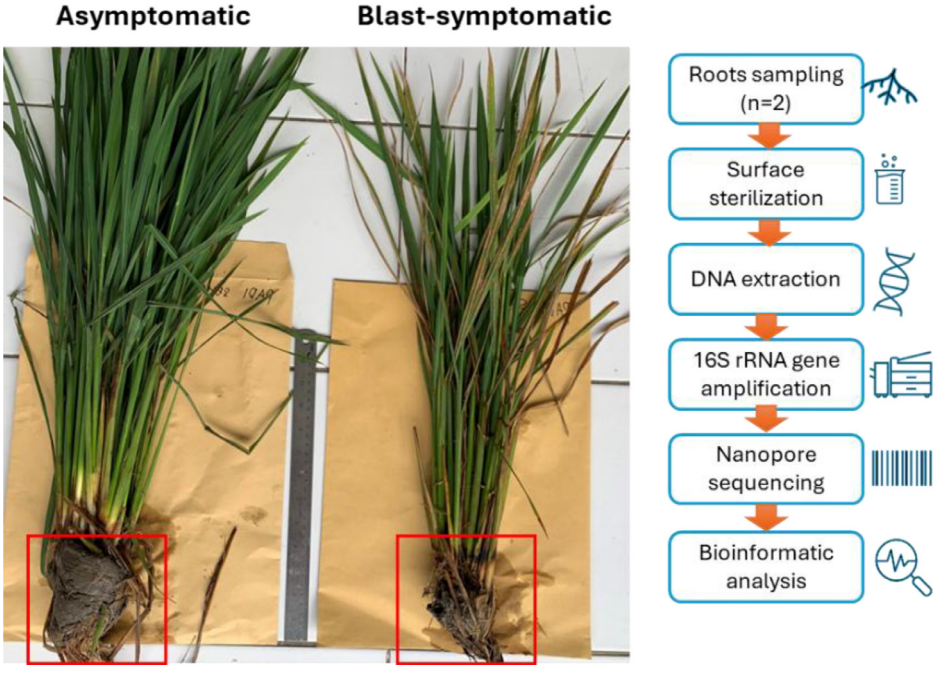

Fig. 4. Overall method of metabarcoding data collection. The red rectangles indicate the representative of root samples before surface sterilization.

4.3. Data analysis

The resulting data were analysed and visualised using two bioinformatics platforms: EPI2ME [13] and MicrobiomeAnalyst v.2.0 (https://www.microbiomeanalyst.ca/) [14]. The workflow commenced with the generation of a feature table using EPI2ME, followed by downstream analysis

and visualisation via MicrobiomeAnalyst. Additionally, the taxonomic composition of the microbial communities was visualised using Krona plots generated with Krona Tools.

Limitations

The scope of the present study is strictly confined to the characterization of endophytic bacterial communities. Accordingly, this investigation does not encompass the experimental validation of the specific impacts exerted by beneficial isolates on plant growth, yield, or disease resistance. Future research initiatives are designated to elucidate these functional attributes and verify the physiological contributions of the identified taxa

Ethics Statement

Not applicable.

Credit Author Statement

Yasir Sidiq: Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing, Funding acquisition**. Triastuti Rahayu:** Conceptualization, Formal analysis, Validation, Funding acquisition. **Peni Indrayudha:** Conceptualization, Resources, Funding acquisition**. Erma Musbita Tyastuti:** Investigation, Data curation**. Azmi Zaki Waliudin Althaf:** Data Analysis**. Banuwati Kartika Sari:** Writing – original draft).

Data availability

PRJNA992961 (Original data) (NCBI)

Acknowledgments

We express our gratitude to Universitas Muhammadiyah Surakarta for funding this study through the Hibah Integrasi Tridharma (HIT) programme (Contract Number 775/A.3-III/FKIP/V/2022). We also extend our thanks to Ms Riya Boita and Ms Diyah Ramadani for their technical laboratory support.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] T.G. Ercole, D.R. Bonotto, M. Hungria, V.M. Kava, L.V. Galli, The role of endophytic bacteria in enhancing plant growth and health for sustainable agriculture, Antonie Van Leeuwenhoek 118 (2025) 88, doi:10.1007/s10482-025-02100-0.
[2] T. Rahayu, Y.A. Purwestri, S. Subandiyah, A. Suparmin, D. Widianto, Exploration of core endophytic bacteria from different organs of diploid *Musa balbisiana* and triploid *Musa acuminata*, Agric. Nat. Resour. 55 (5) (2021) 787–794, doi:10.34044/j.anres.2021.55.5.09.

[3] U. Hofer, The majority is uncultured, Nat. Rev. Microbiol. 16 (2018) 716–717, doi:10.1038/s41579-018-0097-x.
[4] D.M. Tran, T.H. Nguyen, Endophytic bacterial dataset of the Cavendish banana grown in Dak Lak province of Vietnam using 16S rRNA gene metabarcoding, Data Br. 52 (2024) 109863, doi:10.1016/j.dib.2023.109863.
[5] D.M. Tran, T.H. Nguyen, 16S rRNA metagenomic dataset on endophytic bacterial community of the cashew plant (Anacardium occidentale L.) grown in Dak Lak province of Vietnam, Data Br. 52 (2024) 110039, doi:10.1016/j.dib.2024.110039.
[6] D.M. Tran, T.H. Nguyen, T.O. Huynh, T.O. Do, Q.V. Nguyen, A.D. Nguyen, Analysis of endophytic microbiome dataset from roots of black pepper (Piper nigrum L.) cultivated in the Central Highlands region, Vietnam using 16S rRNA gene metagenomic next-generation sequencing, Data Br. 42 (2022) 108108, doi:10.1016/j.dib.2022.108108.
[7] T.H. Nguyen, D.M. Tran, Root endophytic microbiome dataset of sugarcane (Saccharum officinarum L.) cultivated in the Central Highlands, Vietnam, established by the 16S rRNA metagenomics, Data Br. 48 (2023) 109103, doi:10.1016/j.dib.2023.109103.
[8] T. Rahayu, E.M. Tyastuti, A. Ambarwati, L. Agustina, N.A. Setiyadi, N. Jamil, Y. Sidiq, Metagenomic data of bacterial 16S rRNA in the cemetery soil samples in Surakarta City, Indonesia, Data Br. 52 (2024) 109963, doi:10.1016/j.dib.2023.109963.
[9] A.G.M.P. Hardianto, T. Rahayu, E.M. Tyastuti, L. Agustina, N.A. Setiyadi, Y. Sidiq, Screening and identification of lipolytic potential of actinobacteria from cemetery in Bonoloyo, Surakarta, Central Java, Indonesia, Biodiversitas 26 (2025) 2786–2793, doi:10.13057/biodiv/d260623.
[10] R.R. Wick, L.M. Judd, K.E. Holt, Performance of neural network basecalling tools for Oxford Nanopore sequencing, Genome Biol. 20 (2019) 129, doi:10.1186/s13059-019-1727-y.
[11] W. De Coster, S. D'Hert, D.T. Schultz, M. Cruts, C. Van Broeckhoven, NanoPack: visualizing and processing long-read sequencing data, Bioinformatics 34 (15) (2018) 2666–2669, doi:10.1093/bioinformatics/bty149.
[12] A.B. Nygaard, H.S. Tunsjø, R. Meisal, C. Charnock, A preliminary study on the potential of Nanopore MinION and Illumina MiSeq 16S rRNA gene sequencing to characterize building-dust microbiomes, Sci. Rep. 10 (2020) 3209, doi:10.1038/s41598-020-59771-0.
[13] EPI2ME Labs 23.02-01 Release, EPI2ME Labs blog (2023). Available at: https://epi2me.nanoporetech.com/epi2me-labs-23.02.01-release/.
[14] Y. Lu, G. Zhou, J. Ewald, Z. Pang, T. Shiri, J. Xia, MicrobiomeAnalyst 2.0: comprehensive statistical, functional and integrative analysis of microbiome data, Nucleic Acids Res. 51 (W1) (2023) W310–W318, doi:10.1093/nar/gkad407.