Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Metagenomic Insights into Lignocellulose-Degrading Microbiota in High-Temperature (75°C) Chicken Manure and Straw Compost via Nanopore Sequencing

Tigran Soghomonyan
Armbiotechnology SPC
Lev Khoyetsyan
Armbiotechnology SPC
Artur Hambardzumyan
arthambardzumyan@gmail.com
Armbiotechnology SPC
Oleg Reva
University of Pretoria
Marina Paronyan
Armbiotechnology SPC
Karine Dyukova
Armbiotechnology SPC
Marieta Izmailyan
Armbiotechnology SPC
Anichka Hovsepyan
Armbiotechnology SPC
Anna Mkhitaryan
Armbiotechnology SPC

Research Article

Keywords: Thermophilic microbial consortium, Chicken manure and straw compost, Metagenomics, Lignocellulose-degrading enzymes, Heterologous gene expression

Posted Date: June 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9779007/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

The discovery of robust biocatalysts from extreme environments is essential for advancing sustainable second-generation biofuel production. This study investigates the lignocellulose-degrading potential of a thermophilic microbial consortium derived from industrial chicken manure and straw compost maintained at 75°C. Using Oxford Nanopore Technologies (ONT) long-read sequencing, we performed a de novo metagenome assembly, resulting in a 25.1 Mb functional metagenome containing 40,763 coding sequences. Taxonomic analysis revealed a heavy dominance of the superkingdom Bacteria (99.41%), primarily represented by the phyla Deinococcota (73.18%), Pseudomonadota (19.17%), and Bacillota (4.92%), indicating a specialized community shaped by intense thermal selective pressure.

Metagenomic mining identified five candidate genes encoding lignocellulose-degrading enzymes, including two xylanases and two peroxidases with relatively low amino acid sequence homology (75–82%) to known functional sequences. These genes were successfully expressed heterologously in *Bacillus subtilis* RIK1285. Recombinant strains demonstrated significant functional enhancement; the volumetric xylanase activity increased more than three-fold compared to the host strain, while the recombinant peroxidases exhibited a 20-fold increase in lignin-decomposition rates at pH 7.2. These findings validate the efficacy of long-read metagenomic pipelines in uncovering novel, highly active biocatalysts. The discovered enzymes, characterized by their exceptional stability and activity under thermophilic and acidic-to-neutral conditions, hold substantial promise for the industrial valorization of recalcitrant agricultural biomass.

Key points

• 25.1 Mb metagenome (ONT) – 40,763 coding sequences

• Bacteria 99.41%: Deinococcota 73.18%, Pseudomonadota 19.17%, Bacillota 4.92%

• Expressed in B. subtilis: >3× xylanase and ~ 20× peroxidase activity; thermotolerant

Introduction

Functional metagenomics, through the integration of next-generation sequencing technologies, bioinformatics, and genetic engineering, enables the discovery of enzymes with novel technological characteristics, even within uncultivable communities. These enzymes can be optimized for technological use and applied in various processes to produce value-added products (Robinson et al. 2021; Hogg et al. 2024).

The global abundance of lignocellulose-rich raw materials underscores the significant potential of leveraging functional metagenomes to identify enzymes for their efficient degradation (Beloqui et al. 2010; Dougherty et al. 2012). Recombinant metagenomic lignocellulose-decomposing enzymes are instrumental in dismantling these complex biomaterials, serving as biocatalysts for the transformation of plant biomass into sustainable high-value products such as biofuels and bioplastics (Dougherty et al. 2012; Ellilä et al.

2019). This biotechnological strategy has successfully yielded a broad spectrum of novel enzymes from diverse environments as shown in Table 1.

Table 1
Industrial enzymes derived from environmental microbial consortia

Enzyme	Source of isolation
Glycosyl hydrolases (GH) displaying unique substrate specificities such as dual xylanase and cellulase activities.	Calf rumen (Ferrer et al. 2012), compost (Sae-Lee and Boonmee 2014), earthworm casts (Beloqui et al. 2010), and rainforest decaying logs (Nyathi et al. 2024).
Novel hemicellulases and β-glucosidases with improved industrial properties, such as high glucose tolerance and strong transglycosylation activity.	Compost (Dougherty et al. 2012; Uchiyama et al. 2013).
Thermostable bacterial xylanases.	Lignocellulose-enriched compost (Ellilä et al. 2019).
GH family 9 endoglucanases with varied modular architectures	Compost (Aymé et al. 2021).
Specialized lignin-modifying enzymes, including lignin peroxidases, manganese peroxidases, and laccases, essential for overcoming the recalcitrance of the lignin barrier	Forest soils with leaf litter and decaying wood ecosystems (Pollegioni et al. 2015).

Enzymes derived from thermophilic microbial communities are particularly promising due to their exceptional thermal and operational stability, allowing industrial bioprocesses to function at temperatures reaching 80–90°C, where standard fungal preparations are typically deactivated (Gladden et al. 2011; Sato et al. 2017; Busch et al. 2021). Metagenomic mining of diverse niches, ranging from hot spring soils to switch grass-adapted consortia and composting residues, has uncovered highly novel biocatalysts that may share less than 50% amino acid sequence similarity with known protein sequences (Xia et al. 2013; Colombo et al. 2016; Sato et al. 2017; Reichart et al. 2021). These enzymes offer unique advantages for industrial applications, including remarkable tolerance to end-product inhibition (such as high xylose concentrations) and resilience to pretreatment chemicals like ionic liquids (Gladden et al. 2011; Ndata et al. 2021). Furthermore, the identification of specialized metabolic structures, such as thermo-stable cellulosome-based systems, allows for more efficient and intimate association with recalcitrant biomass (Xia et al. 2013). These enhanced capabilities significantly improve process efficiency and control, providing critical tools for the cost-effective production of sustainable second-generation biofuels (Xia et al. 2013; Colombo et al. 2016; Reichart et al. 2021).

Microbial genomes derived from metagenomic data can be assembled using short-read sequencing technologies, such as Illumina. However, the contiguity of genomes assembled from such metagenomes is often limited by challenges in resolving repetitive genomic elements. Accurate decoding of these repetitive regions is crucial for reconstructing complete and correct genome structures. In this context, integrating long-read sequencing technologies, such as Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio), with short-read sequencing methods is indispensable (Rhoads and Au 2015; Moss et al. 2020). Currently, we are enhancing our expertise in ONT sequencing, particularly because of its balance of

accuracy and affordability. Notably, ONT sequencing allows for the assembly of nearly complete bacterial genomes directly from metagenomic reads, eliminating the need for additional short-read data (Sereika et al. 2022).

This study aims to sequence the metagenomic DNA of a thermophilic microbial consortium (75°C) derived from compost produced under industrial conditions using chicken manure and straw. The objectives are to assess the biodiversity of the metagenome, process the raw reads through assembly, annotation, and binning, and identify potential novel lignocellulose-degrading enzymes.

Materials and Methods

2.1. Materials

All chemicals used in this work (analytical grade) were purchased from Sigma-Aldrich (St. Louis, MO, USA). RNase A (DNase- and protease-free) was obtained from Thermo Fisher Scientific (Waltham, MA, USA). Proteinase K, lysozyme, lysostaphin, and lyticase were purchased from Sigma-Aldrich (St. Louis, MO, USA).

2.2. Compost Sampling, microbial biomass isolation, and cryopreservation

Compost sampling was conducted from the second-phase compost (75°C) at the composting department of the "Spitak Glkhark" mushroom (*Agaricus bisporus*) production facility near Jraber village, Armenia (40°20'16.7"N, 44°39'19.3"E). This compost was based on straw and chicken manure and was prepared in an industrial manner as follows:

The composting material was stored in a concrete bunker (dimensions: 50 × 5 × 3 m) with bottom aeration. Every 4–5 days, the material was removed to a concrete square, watered with a 4–5% chicken manure suspension, and transferred to another bunker of the same dimensions. This process was repeated until the compost reached an optimal carbon-to-nitrogen (C:N) ratio (typically 25:1 to 30:1) and moisture content (50–60%). The composting process was then continued for an additional 1–1.5 months, or until the temperature decreased to match the surrounding ambient temperature.

To mine genes encoding lignocellulose-degrading thermophilic enzymes, samples were taken from the second-phase thermophilic compost at a temperature of 75°C. The samples were collected from the center of the bulk at a depth of 0.5 m, amounting to 2 kg of compost.

The microbial biomass of the compost-inhabiting microbial consortia was prepared using a modified method of Vishan et al. (Vishan et al. 2017). For this, 20 g of wet compost was placed in each of 20 Erlenmeyer flasks (500 mL capacity), and 180 mL of sterile 0.85% (w/v) NaCl solution was added to each flask. Mechanical mixing was performed at 150 rpm for 16 h at 25°C. All suspensions were combined into a 5-liter tank, allowed to settle at room temperature for 4 hours, and decanted from the settled material. The microbial biomass was then centrifuged, distributed into 0.6–0.8 g portions in 2 mL Eppendorf tubes as wet biomass, frozen in liquid nitrogen, and stored at − 80°C until use.

2.3. DNA extraction and metagenomic sequencing

To isolate high-molecular-weight, high-purity DNA from microbial biomass collected from compost for long-read Nanopore sequencing, we developed a comprehensive procedure. This included the preliminary treatment of the metagenomic biomass using a combination of modified methods derived from Bag et al. (Bag et al. 2016) and Busch et al. (Busch et al. 2021), followed by final purification with the Monarch® Genomic DNA Purification Kit (New England Biolabs, UK). The procedure is as follows:

The contents of the cryopreserved microbial biomass tube were suspended in a 10 mL plastic vial with 2 mL of a solution containing 50 mM Tris-HCl buffer (pH 8.0), 20 mM EDTA, 20 mg lysozyme, 2 µL lysostaphin, and 20 µL lyticase. The mixture was then incubated for 1 hour at 37°C. After incubation, 2.5 mL of 4 M guanidinium thiocyanate solution was added to the suspension and gently mixed for 45 seconds at room temperature. Next, 3.5 mL of 10% lauroyl sarcosine solution was added, and the mixture was incubated for 10 minutes in a vortex mixer at 300 rpm and 37°C, followed by static incubation for 1 hour at 70°C. At the end of this treatment, the suspension was subjected to bead-beating with glass beads (2 × 0.5 minutes) using a vortex machine, with 300 mg of glass beads per 750 µL of solution. In the next step, 15 mg of polyvinylpyrrolidone (20 kDa) was added per 750 µL of the suspension, gently mixed, and the glass beads were removed by centrifugation at 18,000 × g using an Eppendorf centrifuge. The pellet from each 750 µL of suspension was washed with 500 µL of a solution containing 50 mM Tris-HCl buffer (pH 8.0), 20 mM EDTA, 100 mM NaCl, and 1% polyvinylpyrrolidone. After centrifugation, the supernatant was combined with the supernatant obtained in the previous step. To each 950 µL of combined supernatant, 50 µL of 5 M potassium acetate buffer (pH 5.2) was added in 2 mL Eppendorf tubes and gently mixed. Next, 1 mL of a phenol:chloroform:isoamyl alcohol mixture (25:24:1) was added, mixed to form an emulsion, and centrifuged at 18,000 × g for 5 minutes. The upper phase was carefully transferred to a new Eppendorf tube, an equal volume of isopropanol was added, and the mixture was gently mixed. After being stored in a refrigerator for at least 1 hour at -20°C, it was centrifuged at 18,000 × g for 5 minutes.

The precipitate obtained in the previous step was further processed using the Monarch® Genomic DNA Purification Kit as follows: The precipitate was dissolved in 100 µL of PBS buffer, followed by the addition of 1 µL of proteinase K and 3 µL of RNase A. After gentle mixing, 100 µL of Monarch cell lysis buffer was added, and the mixture was incubated for 5 minutes in a vortex mixer at 1400 rpm and 56°C. The spin column binding and washing steps were performed according to the manufacturer's instructions. In the final elution step, 100 µL of 60°C Monarch genomic DNA elution buffer was added, and DNA was eluted by centrifugation at 12,000 g for 1 minute. After analysis, the metagenomic DNA was frozen using liquid nitrogen and stored at -80°C until use.

Metagenomic sequencing was performed according to the recommendations of Oxford Nanopore Technologies (UK). ONT's proprietary MinKNOW software was utilized for raw data collection and basecalling. DNA library preparation, as well as priming and loading of the Spot ON MinION R9.4.1 flow cell, was conducted following the ONT gDNA-sqk-lsk109 protocol (GDE_9063_v109_revAP_25May2022.pdf) available at https://nanoporetech.com/es/document/ gDNA-sqk-lsk109?format=versions.

2.4. Study of the biodiversity of the obtained metagenome

The EPI2ME Desktop (version 5.2.3), developed by Metrichor Ltd., a subsidiary of Oxford Nanopore Technologies, was used for taxonomic characterization of raw data in FASTQ format. The EPI2ME Desktop workflow “Metagenomics” was employed for this purpose. The software is available at: https://nanoporetech.com/document/epi2me?format=versions.

2.5. Assembly, annotation, polishing, and binning of the obtained metagenome

Metagenome assembly was performed using Flye (version 2.9.5-b1801), a **de novo** genome assembler designed for long-read sequencing data. It is optimized for assembling high-error-rate reads into high-quality, polished contiguous sequences (contigs) (Kolmogorov et al. 2019; Kolmogorov et al. 2020). The software is available at: https://github.com/mikolmogorov/Flye/releases/tag/2.9.5.

Metagenome annotation was performed using Prokka (version 1.14.6), a prokaryotic genome annotator that provides high-quality, standardized annotations (Seemann 2014). The software is available at: https://github.com/tseemann/prokka/releases/tag/v1.14.6.

Metagenome assemblies were polished using ONT’s Medaka (version 2.0.1). Medaka improves base-level accuracy by correcting sequencing errors, making it particularly useful for metagenome binning, where high-quality contigs are essential for accurate taxonomic classification. The software is available at: https://github.com/nanoporetech/medaka.

Medaka-polished assemblies *were binned* using MaxBin2 (version 2.2.7_1) (Wu et al. 2016). The software is available at: https://github.com/brewsci/homebrew-bio/pkgs/container/bio%2Fmaxbin2.

The MaxBin2-generated bins *were taxonomically characterized* using the online Kaiju server (https://github.com/bioinformatics-centre/kaiju), whose services were discontinued on 2024-07-31 (Menzel et al. 2016). Version 1.10.1 (released 2024-03-03) was used for the analyses.

2.6. Mining novel lignocellulose-degrading enzymes in the obtained metagenome

Enzyme mining

The PROKKA_03062024.gbk file from Mendeley Data was manually searched for the terms “glucanase,” “xylanase,” and “peroxidase.” Coding sequences (CDS) that met the selection criteria, appropriate translation length, > 70% homology to known functional enzymes, and presence of the catalytic domain in InterPro analyses, were selected for primer design and gene cloning.

PCR amplification: Putative β-glucanase, xylanase, and peroxidase genes were amplified by polymerase chain reaction (PCR). Thermal cycling conditions were: initial denaturation at 95°C for 2 min; 30 cycles of 95°C for 30 s (denaturation), 57°C for 30 s (annealing), and 72°C for 2 min (extension); and a final extension at 72°C for 5 min.

DNA electrophoresis

PCR products were resolved on 0.9% agarose gels prepared in 40 mM Tris–acetate–EDTA (TAE) buffer, pH 8.0 (Agarose I, VWR tablets). Electrophoresis was performed at 100 V for ~ 35 min. DNA bands were visualized with GelRed® nucleic acid stain (Millipore), and fragment sizes were estimated using the TriDye™ 1 kb Plus DNA ladder (New England Biolabs).

Cloning: Cloning was performed using Gibson assembly (Gibson et al. 2009). The pBE-S shuttle vector (Takara Bio) was linearized by double digestion with NdeI and XbaI (New England Biolabs) according to the manufacturer's protocol. Amplified gene fragments were assembled into the linearized vector using NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs) at 50°C following the manufacturer's instructions; reactions were carried out in a 10 µL final volume with insert:vector ratios as recommended.

The primer sequences used in this study are summarized in Table 2. Restriction sites are highlighted in gray. These sites are preceded by vector-homology motifs and followed by target-gene homology motifs.

Table 2
Primers used in this study

Gene	Primer name	Primer sequence (5'–3')
β-Glucanase_28564	Gluc_28564_F	GCCGGTGCACATATGGGCGGTCACGGCTGGG
β-Glucanase_28564	Gluc_28564_R	ATGGTGATGTCTAGACTACTCCACCCATCGGTACACG
Xylanase_14874	Xyl_14874_F	GCCGGTGCACATATGGGAAAGTCTAGAACCCAATCGGGG
Xylanase_14874	Xyl_14874_R	ATGGTGATGTCTAGACTACTTTGCACCCAGGTGCAAAAG
Xylanase_34975	Xyl_34975_F	GCCGGTGCACATATGGCCGTTACGTTCAACATGGCC
Xylanase_34975	Xyl_34975_R	ATGGTGATGTCTAGATCAGAAAATCCTCCATGAACGAAGC
Peroxidase_05038	Pox_05038_F	GCCGGTGCACATATGACAGCTCAATCTGTAATCTTGCCATTAC
Peroxidase_05038	Pox_05038_R	ATGGTGATGTCTAGATTATAAATTATCCTTCAGCACTGTTTCCAGC
Peroxidase_20739	Pox_20739_F	GCCGGTGCACATATGTTTTGGCTTTTGCGTCCAACACC
Peroxidase_20739	Pox_20739_R	ATGGTGATGTCTAGATTACACTTTTAAGTAGTCGGCCAATTTGTC

Culture conditions

The low-protease strain *Bacillus subtilis* RIK1285 (Takara Bio) was used as the host for heterologous expression of recombinant enzymes. *Escherichia coli* Top10 (Invitrogen) was used for plasmid propagation and cloning. Bacterial strains were maintained on Luria–Bertani (LB) agar plates at 4°C or preserved as glycerol stocks in LB medium containing 50% (v/v) glycerol at − 80°C. For routine cultivation, cultures were grown in LB medium (1% peptone, 0.5% yeast extract, 0.5% NaCl) on a rotary shaker at 150 rpm. Incubation temperatures were 37°C for *E. coli* and 33°C for *B. subtilis*. When required, media were supplemented with ampicillin (100 µg/mL) for *E. coli* or kanamycin (10 µg/mL) for *B. subtilis*. Overnight cultures grown under these conditions were used for plasmid isolation and molecular cloning.

Transformation

Assembly mixtures were transformed into chemically competent *Escherichia coli* Top10 cells by heat shock. Transformants were selected on LB agar plates containing ampicillin (100 µg/mL). Positive colonies were screened by colony PCR, and plasmid DNA was isolated using the HiGene™ Plasmid Mini Prep Kit according to the manufacturer's instructions. Purified plasmids were stored at − 20°C until further use in transformation and expression experiments.

Competent *Bacillus subtilis* RIK1285 cells were transformed with recombinant pBE-S–based expression plasmids constructed in this study. Transformations followed the Takara Bio protocol for *Bacillus* transformation with modifications. Briefly, cultures were grown under nutrient-limiting conditions to induce the physiological state associated with sporulation and the development of natural competence. Cells harvested at the competence stage were used for plasmid uptake (Soghomonyan et al. 2024).

Fermentation: Fermentation experiments with recombinant *Bacillus subtilis* RIK1285 were conducted in a defined production medium with the following components (%, w/v): glycerol 2.0, yeast extract 0.5, peptone 1.0, $(NH_4)_2SO_4$ 0.3, NH_4Cl 0.2, KH_2PO_4 0.5, and $MgSO_4$ 0.025. The medium was supplemented with L-lysine and L-tryptophan (40 mg/L each) and kanamycin (10 mg/ L) for plasmid maintenance. The initial pH was adjusted to 7.2 prior to sterilization.

For inoculum preparation, an overnight culture was grown in Luria–Bertani (LB) medium at 33°C on a rotary shaker at 180 rpm. The seed culture was used to inoculate the fermentation medium at a 5% (v/v) inoculation ratio. Shake-flask fermentations were performed in 500 mL cotton-plugged Erlenmeyer flasks containing 20 mL of fermentation medium. Cultures were incubated at 33°C with agitation at 220 rpm to ensure adequate aeration and homogeneous growth.

Enzyme activity

Xylanase activity was assayed in a 0.2 mL reaction mixture containing 1% (w/v) pre-boiled beechwood xylan (Megazyme, Neogen), 50 mM acetate buffer (pH 6.0) or 50 mM phosphate buffer (pH 7.2), and appropriately diluted enzyme solution, following a modified method of Saha (Saha 2001). After incubation at 60°C, reducing sugars released were measured by the 3,5-dinitrosalicylic acid (DNS) method (Miller, G.L. 1959). One unit (U) of xylanase activity was defined as the amount of enzyme that releases 1 µmol reducing sugar per minute under the specified conditions. For qualitative estimation of xylanase activity, 0.1% AZCL-xylan (Megazyme, Neogen) was incorporated in 1% agarose gels prepared in 50 mM acetate buffer (pH 6.0) (Agarose I™, VWR). Photographs were taken after 1 h incubation at 37°C.

Peroxidase activity was determined by the decolorization of 1,4-phenylenediamine (p-PDA) at 40°C in a 0.7 mL reaction mixture containing 1 mM p-PDA, 0.25 mM H_2O_2 and 50 mM acetate buffer (pH 4.5), with appropriately diluted enzyme, following a modified method of Zhang et al. (Zhang et al. 2017). The decrease in absorbance was measured at 284 nm using an extinction coefficient of 2410 M^{-1} cm^{-1}. One unit (U) of peroxidase activity is the amount of enzyme that decolorizes 1 µmol of p-PDA per minute under the specified conditions.

Lignin-degrading activity of the obtained peroxidases was determined by monitoring the increase in phenolic concentration at 280 nm at 40°C in a 1-mL reaction mixture containing 0.5% lignin, 1 mM H_2O_2, and 50 mM acetate buffer (pH 4.5) or phosphate buffer (pH 7.2), with appropriately diluted enzyme, following a modified method of Silva et al. (Silva et al. 2023).

Protein concentration was estimated by the Groves and Davis method (Groves et al. 1968).

Statistical Analysis

Statistical analysis results are presented as mean ± standard deviation of three independent experiments. Calculations were performed using Microsoft Office Excel 2016. p-values were determined using IBM SPSS Statistics 16.0 (IBM Corp., Armonk, NY, USA). Differences were considered statistically significant at $p < 0.05$.

Results

3.1. Metagenomic DNA extraction

Metagenomic DNA was prepared using the procedure outlined in Section 2.3 from the frozen compost biomass obtained via the method described in Section 2.2. The main characteristics of the obtained DNA are presented in Supplementary Fig. S1.

The concentration of the obtained DNA sample was 108.8 μg/mL, with A_{260}/A_{230} and A_{260}/A_{280} ratios of 2.56 and 2.47, respectively. The presented data clearly indicate that the main component of the DNA sample obtained using the described procedure exceeds 20 kbp, which is crucial for long-read nanopore sequencing.

3.2. Sequencing of metagenomic DNA

Metagenomic DNA was sequenced following the procedure outlined in Section 2.3. The main sequencing parameters generated by MinKNOW are presented in Supplementary Fig. S2. The obtained data indicate a relatively small metagenome – 1.82 Gb of estimated bases, with 559,960 generated reads and an estimated mean read length of 3.66 kb. Of these, 1.09 Gb of estimated bases met the PHRED quality score threshold of 9, and this data was used for subsequent metagenomic analyses.

3.3. Study of the biodiversity of the obtained metagenome

A total of 328,920 DNA reads generated by ONT sequencing were analyzed using EPI2ME Desktop. The average length of reads was 3,269 bp in range from 1,000 to 100,000 bases. After quality filtering in EPI2ME Desktop, 317,393 reads passed the quality threshold and were retained for downstream analysis. Taxonomic profiles were subsequently filtered using the abundance_threshold parameter, which excludes low-abundance taxa below a predefined relative abundance cutoff. Of the total reads, 83,893 (25.51%) were not classified by the software. The classified reads at the superkingdom level are presented in Table 3.

Table 3
Distribution of the classified high quality reads by superkingdoms

Superkingdom	Reads (fastq_pass_M2)	Reads, (%)
Bacteria	232,119	99.41
Eukaryota	1,264	0.54
Viruses	117	0.05

Of the 317,393 reads retained after EPI2ME Desktop filtering, the majority were assigned to the superkingdom Bacteria (99.41%), followed by Eukaryota (0.54%) and Viruses (0.05%). The distribution of the main bacterial representatives at the order level is shown schematically in Fig. 1.

Of the 317,393 high-quality reads assigned to the superkingdom Bacteria, the reads are distributed among three groups: Thermotogati (170,483), Pseudomonadati (47,834), and Bacillati (13,802). Phylum-level data are presented in Fig. 1, Supplementary Table S1, and in the corresponding sheets of *wf-metagenomics-counts.xlsx* and *wf-metagenomics-rarefied.xlsx* deposited in Mendeley Data. The major phyla include Deinococcota (associated with Thermotogati), Pseudomonadota (associated with Pseudomonadati), and Bacillota (associated with Bacillati), as shown in Fig. 1; minor phyla for each group are listed in Table S1 and the accompanying Excel files.

At the class level, the metagenome comprises 21 taxa; at the genus level, 127 taxa; and at the species level, 322 taxa (including Unknown, Eukaryota, and Viruses). Full details are provided in the Excel files deposited in Mendeley Data. Species richness is also shown in the file Fig_species.jpg in the Mendeley Data collection.

3.4. Metagenome assembly, annotation, polishing, and binning

The metagenome assembly and annotation are presented in Table 4. The Flye assembly of the metagenome revealed 935 contigs with a total length of 25,063,188 nucleotides and a mean coverage of 38. Among these, contigs 329, 1093, 57, 1035, 785, 627, 509, 632, 1049, 64, 71, 12, 493, and 5 were circular, with lengths ranging from 20,248 to 1,823 nucleotides. Contig information is provided in the assembly_info.txt file on Mendeley Data. Additional assembly details and sequences are available in the flye.log and assembly.fasta files on Mendeley Data.

Table 4
Metagenome assembly and annotation summary

Flye assembled	Prokka annotated
Total length: 25,063,188	Contigs: 824
Fragments: 935	Bases: 24,859,457
Fragments N50: 71,931	CDS: 40,763
Largest frg: 973,877	rRNA: 73
Scaffolds: 0	tRNA: 432
Mean coverage: 38	Repeat_region: 31
	tmRNA: 6

Prokka annotation was performed after removing contigs smaller than 3,000 bp. The remaining 824 contigs contained a total of 24,859,457 bp. The summary of the Prokka annotation is presented in Table 4. Of the 40,763 CDS, nearly 75% were annotated by Prokka, while the remaining CDS were hypothetical proteins. The numbers of annotated rRNA, tRNA, and tmRNA genes, as well as repeat regions, are provided in Table 4. The files *PROKKA_03062024.gbk*, *PROKKA_03062024.faa*, and *PROKKA_03062024.ffn*, containing core information on the M2 metagenome annotation, are available on Mendeley Data.

For binning the metagenome data, Medaka polishing was performed using the metagenome raw reads and the Flye assembly in FASTA format. The resulting consensus.fasta file is available on Mendeley Data. Binning with MaxBin2 was conducted based on the information in the polished consensus.fasta file, resulting in four bins, which are characterized in Table 5.

Table 5
M2 metagenome bins characteristics

Bin name	Completeness, %	Genome size	GC content, %
Maxbin2_medaka_bin.001.fasta	86.9	5566583	40.8
Maxbin2_medaka_bin.002.fasta	62.6	4411211	46.1
Maxbin2_medaka_bin.003.fasta	81.3	9962125	65.7
Maxbin2_medaka_bin.004.fasta	67.3	4934814	67.7

The obtained files, *Maxbin2_medaka_bin.001.fasta*, *Maxbin2_medaka_bin.002.fasta*, *Maxbin2_medaka_bin.003.fasta*, and *Maxbin2_medaka_bin.004.fasta*, are available on Mendeley Data. These bins comprise more than 99% of the information in the polished M2 metagenome. The remaining unbinned 26 contigs are collected in the Maxbin2_medaka_bin.noclass file, which is available on Mendeley Data.

The taxonomic composition of bins 001–004, derived from the Kaiju web server, is presented in Supplementary Fig. S3–S6, respectively. The data indicate that bins 001 and 004 are relatively uniform,

being dominated by Bacillota (formerly Firmicutes) and Deinococcota (formerly Deinococcus–Thermus), respectively, while bins 002 and 003 contain more diverse taxa. The main constituents of bin 002 are Bacillota and Pseudomonadota (formerly Proteobacteria). The main constituents of bin 003 are Bacteroidota (formerly Bacteroidetes), Bacillota, and Pseudomonadota.

3.5. Estimating the potential for novel lignocellulose-degrading enzymes in the obtained metagenome

The following lignocellulose-degrading enzymes were identified in the M2 metagenome: β-glucanase_28564 (NGJGEFEI_28564), xylanase_14874 (NGJGEFEI_14874), xylanase_34975 (NGJGEFEI_34975), peroxidase_05038 (NGJGEFEI_05038) and peroxidase_20739 (NGJGEFEI_20739). Their relatedness to known enzymes based on NCBI protein BLAST is presented in Table 6.

Table 6
Alignment of lignocellulose-degrading enzymes identified in the M2 metagenome to known proteins

Mined Enzyme (Inferred Function)	**Annotated CDS (Locus Tag)**	**Top BLAST Hit (Description)**	**BLAST Homology, %**	**Reference**
β-glucanase_28564	NGJGEFEI_28564	1,3 – 1,4-β-glucanase of *Rhodothermus marinus*	93.75	(Spilliaert et al. 1994)
Xylanase_14874	NGJGEFEI_14874	Endo-1,4-β-xylanase of *Deinococcota bacterium*	100.00	Unpublished
		Endo-1,4-β-xylanase of *Thermus brockianus*	75.56	(Schäfers et al. 2017)
Xylanase_34975	NGJGEFEI_34975	Endo-1,4-β-xylanase of *Rhodothermus marinus*	82.51	(Chu et al. 2017)
Peroxidase_05038	NGJGEFEI_05038	Dyp-type peroxidase of *Acinetobacter indicus*	98.03	(Colpa et al. 2014; Sugano and Yoshida 2021)
Peroxidase_20739	NGJGEFEI_20739	Heme-dependent peroxidase of *Bacilli bacterium*	99.44	(Campanaro et al. 2020)

The amino acid and nucleotide sequences of all five proteins from Table 6 are presented in Supplementary Fig. S7–S11 (Mendeley Data).

Primer pairs were designed (see Section 2.6, Materials and Methods) to amplify β-Glucanase_28564, Xylanase_14874, Xylanase_34975, Peroxidase_05038, and Peroxidase_20739 genes for Gibson assembly

cloning into the pBE-S shuttle vector for extracellular expression. All genes except β-Glucanase_28564 were amplified from M2 metagenomic DNA. The amplified inserts were cloned into pBE-S, propagated in *E. coli* Top10, and subsequently transformed into *B. subtilis* RIK1285 using a specially developed method (Soghomonyan et al. 2024). Colony PCR results for the clones are shown in Supplementary Fig. S12 (Mendeley Data).

Extracellular xylanase activities of the recombinant strains were measured at their previously determined optimal pH values and compared with those of the carrier strain *B. subtilis* RIK1285. The results of enzyme activity measurements in the cultural liquid (CL) after fermentation, as detailed in Table 7 and Fig. 2, were obtained using the methods described in Materials and Methods.

Table 7
Xylanase activities of the recombinant strains

Xylanase strain producer	**pH**	**Volumetric activity in CL, U/mL**	**Specific activity in CL, U/mg**
B. subtilis RIK1285	7.2	14.4 ± 1.3	2.12 ± 0.19
B. subtilis RIK1285_xyl_14874	7.2	46.3 ± 4.1	5.72 ± 0.51
B. subtilis RIK1285	6.0	11.2 ± 0.9	1.64 ± 0.13
B. subtilis RIK1285_xyl_34975	6.0	30.1 ± 2.8	4.07 ± 0.38

* Values are expressed as the mean ± standard deviation of three measurements, with $p < 0.05$.

The results in Table 7 indicate significantly higher extracellular xylanase activity in the recombinant strains compared to the natural activity of the carrier *B. subtilis* RIK 1285 strain. These results were further confirmed by a qualitative assay on AZCL-xylan agar plates (Fig. 2), where the ratio of enzyme activity zone surface areas corresponds to the enzyme activity values presented in Table 7.

Peroxidase activities of metagenome-derived recombinant enzymes were determined by the p-phenylenediamine decolorization assay at pH 4.5, as described in Materials and Methods. Results are presented in Table 8.

Table 8
Peroxidase activities of the recombinant strains

Peroxidase strain producer	**Volumetric activity in CL, U/mL**	**Specific activity in CL, U/mg**
B. subtilis RIK1285	0.06 ± 0.0047	0.0094 ± 0.0007
B. subtilis RIK1285_pox_05038	0.22 ± 0.0190	0.0287 ± 0.0024
B. subtilis RIK1285_pox_20739	0.39 ± 0.0350	0.0508 ± 0.0045

* Values are expressed as the mean ± standard deviation of three measurements, with $p < 0.05$.

As with the peroxidases, extracellular peroxidase activities substantially exceeded the native peroxidase activity of the carrier strain *B. subtilis* RIK1285. This is also evident from lignin-decomposition experiments

(Supplementary Fig. S13), in which lignin breakdown was monitored for 6 h by measuring phenolic production at 280 nm, as described in Materials and Methods. From these experiments it is evident that the reaction proceeds nearly 20-fold faster at pH 7.2 than at pH 4.5.

Discussion

The exploration of thermophilic metagenomes to discover lignocellulose-degrading enzymes with novel technological properties represents a significant scientific challenge (Colombo et al. 2016; Jensen et al. 2018; Santos-Pereira et al. 2023). To this end, we used the MinION Mk1B platform with an R9.4.1 flow cell to sequence the metagenomic DNA of a thermophilic microbial consortium derived from compost made from chicken manure and straw. ONT nanopore sequencing revealed an estimated 1.82 Gb of bases across 559,960 reads; 1.09 Gb of data passing the quality-score threshold were used for subsequent metagenomic analyses. The resulting functional metagenome is relatively small, containing only 40,763 CDSs, but its value lies in the potentially thermophilic nature of the encoded proteins, which are often intrinsically more stable (Rhee et al. 2005; Tirawongsaroj et al. 2008; Lu et al. 2019).

EPI2ME Desktop analysis of the thermophilic microbial consortium (75°C) recovered 317,393 high-quality reads, of which 233,500 were taxonomically classified, showing a heavy dominance of the superkingdom Bacteria (99.41%) alongside minor representations of Eukaryota and Viruses (Fig. 1). The bacterial community was distributed across 10 phyla and was notably dominated by Deinococcota (73.18%), Pseudomonadota (19.17%), and Bacillota (4.92%). The high prevalence of Deinococcota is particularly significant, as members of this phylum are renowned for their resilience to extreme environmental stressors, which aligns with the 75°C industrial composting environment and suggests a specialized community shaped by intense selective pressure for biomass degradation under thermophilic conditions. At lower taxonomic levels, the composition included 18 classes, 124 genera, and 319 species; detailed phylum- to species-level data are provided in the *wf-metagenomics-counts.xlsx* and *wf-metagenomics-rarefied.xlsx* files on Mendeley Data. While the raw counts provide a comprehensive view of the population, the rarefied data, generated through random subsampling to equalize library size, offers a normalized perspective, albeit with the potential removal of low-abundance taxa.

Metagenome binning further elucidated this structure, yielding four distinct bins with varying degrees of completeness. The dominance of Bacillota in bin_001 and Deinococcota in bin_004 suggests these groups are the primary contributors to the metabolic backbone of the consortium. The identification of over 40,000 CDSs within a relatively small metagenome (1.82 Gb) underscores the high coding density and functional specialization of this thermophilic community toward lignocellulose breakdown.

The core objective of this study was the identification and validation of novel lignocellulose-degrading enzymes. We successfully mined and cloned four candidate genes: two xylanases (NGJGEFEI_14874 and NGJGEFEI_34975) and two peroxidases (NGJGEFEI_05038 and NGJGEFEI_20739). Failure to amplify β-Glucanase_28564 likely reflects low template abundance in the metagenomic DNA, primer–template mismatches due to sequence heterogeneity, or assembly errors in the corresponding contig. Future experiments will clarify the cause.

BLAST homology results (Table 5) showed that while some enzymes share high similarity with known sequences (e.g., Peroxidase_20739 at 99.44%), others, such as Xylanase_34975 (82.51%) and Xylanase_14874 (75.56% compared to *Thermus brockianus*), represent potentially novel biocatalysts with distinct properties.

The heterologous expression in *B. subtilis* RIK1285 confirmed the functional viability of these enzymes. The recombinant xylanase strains exhibited significantly higher volumetric and specific activities than the control (Table 6). For example, *B. subtilis* RIK1285_xyl_14874 showed more than a threefold increase in volumetric activity (46.3 U/mL) over the native strain at pH 7.2. Cloning and characterization of novel bacterial xylanases from lignocellulose-enriched compost metagenomic libraries were reported by Ellilä and coauthors (Ellilä et al. 2019).

Similarly, the recombinant peroxidases demonstrated enhanced lignin-degrading potential. A 20-fold increase in lignin-degradation rate at pH 7.2 compared with pH 4.5 suggests these peroxidases are well adapted to the neutral-to-alkaline conditions often found in industrial composting and pulp processing. For comparison, some characteristics of dye-decolorizing peroxidases from *Bacillus subtilis* and *Pseudomonas putida* were reported by Santos et al. (Santos et al. 2014).

The successful isolation of highly active, thermostable enzymes directly from an industrial composting environment validates the efficacy of our ONT nanopore sequencing and binning pipeline. These results demonstrate that even small metagenomic datasets, when derived from specialized niches like high-temperature chicken manure compost, are rich reservoirs for discovering robust biocatalysts. These enzymes hold significant potential for applications in the production of second-generation biofuels and the valorization of agricultural waste, where stability at elevated temperatures is a critical requirement for process efficiency. Similar findings on the industrial importance of mining lignocellulose-degrading enzymes from metagenomes of thermophilic microbial consortia were reported previously (Sato et al. 2017; Busch et al. 2021; Ndata et al. 2021).

Conclusions

Successful Metagenomic Reconstruction

Using Oxford Nanopore (ONT) long-read sequencing, we successfully assembled a 25.1 Mb functional metagenome from a high-temperature (75°C) chicken manure and straw compost, yielding 40,763 coding sequences.

Specialized Microbial Community

The consortium is heavily dominated by the superkingdom Bacteria (99.41%), specifically the phyla Deinococcota (73.18%), Pseudomonadota (19.17%), and Bacillota (4.92%). This composition reflects intense selective pressure toward thermophilic organisms capable of lignocellulose degradation.

Discovery of Novel Biocatalysts

The study identified five potential lignocellulose-degrading enzymes. Notably, xylanases *NGJGEFEI_14874* and *NGJGEFEI_34975* share relatively low homology (75.56% and 82.51%) with known sequences, indicating their potential as novel industrial enzymes.

Enhanced Functional Activity

Heterologous expression in *Bacillus subtilis* RIK1285 confirmed that the mined genes encode highly active enzymes. Recombinant xylanase strains showed over a 3-fold increase in volumetric activity, and recombinant peroxidases exhibited 20-fold faster lignin-decomposition rates at pH 7.2 compared to pH 4.5.

Industrial Applicability

The robust performance of these enzymes at elevated temperatures and acidic-to-neutral pH values demonstrates their high potential for second-generation biofuel production and the biotechnological valorization of agricultural waste.

Declarations

Author's Contributions:

ArH and OR conceived and designed research. ArH, TS,LKh, MP, KD, MI, AnH and AM conducted experiments. All authors analyzed data. ArH, OR, TS and AM wrote the manuscript. All authors read and approved the manuscript.

Acknowledgements and Funding: This research was supported by the Committee of Higher Education and Science of the Republic of Armenia, within the framework of research project № 21AG-2I090.

Data Availability Statement: The supplementary materials (Supplementary_materials.docx), biodiversity estimating information (*Fig_species.jpg*, *wf-metagenomics-counts.xlsx*, *wf-metagenomics-rarefied.xlsx*), binning information (*Maxbin2_medaka_bin.001.fasta*, *Maxbin2_medaka_bin.002.fasta*, *Maxbin2_medaka_bin.003.fasta*, *Maxbin2_medaka_bin.004.fasta*, *Maxbin2_medaka_bin.noclass*) and genome data files (*PROKKA_03062024.faa*, *PROKKA_03062024.ffn*, *PROKKA_03062024.gbk*, *assembly.fasta*, *assembly_info.txt*, *consensus.fasta*, *flye.log*) for this article are available at https://data.mendeley.com/datasets/rs78r9hrzx/3 under the "M2 Metagenome: Chicken manure and straw compost", project of Mendeley Data.

Ethical approval Not applicable. No studies have been conducted on humans or animals.

Competing Interests: The authors have no financial interests or conflicts of interest.

References

1. Aymé L, Hébert A, Henrissat B, Lombard V, Franche N, Perret S, Jourdier E, Heiss-Blanquet S (2021) Characterization of three bacterial glycoside hydrolase family 9 endoglucanases with different modular

architectures isolated from a compost metagenome. Biochim Biophys Acta BBA - Gen Subj 1865(5):129848. https://doi.org/10.1016/j.bbagen.2021.129848

2. Bag S, Saha B, Mehta O, Anbumani D, Kumar N, Dayal M, Pant A, Kumar P, Saxena S, Allin KH, Hansen T, Arumugam M, Vestergaard H, Pedersen O, Pereira V, Abraham P, Tripathi R, Wadhwa N, Bhatnagar S, Prakash VG, Radha V, Anjana RM, Mohan V, Takeda K, Kurakawa T, Nair GB, Das B (2016) An Improved Method for High Quality Metagenomics DNA Extraction from Human and Environmental Samples. Sci Rep 6(1):26775. https://doi.org/10.1038/srep26775
3. Beloqui A, Nechitaylo TY, López-Cortés N, Ghazi A, Guazzaroni M-E, Polaina J, Strittmatter AW, Reva O, Waliczek A, Yakimov MM, Golyshina OV, Ferrer M, Golyshin PN (2010) Diversity of Glycosyl Hydrolases from Cellulose-Depleting Communities Enriched from Casts of Two Earthworm Species. Appl Environ Microbiol 76(17):5934–5946. https://doi.org/10.1128/AEM.00902-10
4. Busch P, Suleiman M, Schäfers C, Antranikian G (2021) A multi-omic screening approach for the discovery of thermoactive glycoside hydrolases. Extremophiles 25(2):101–114. https://doi.org/10.1007/s00792-020-01214-9
5. Campanaro S, Treu L, Rodriguez-R LM, Kovalovszki A, Ziels RM, Maus I, Zhu X, Kougias PG, Basile A, Luo G, Schlüter A, Konstantinidis KT, Angelidaki I (2020) New insights from the biogas microbiome by comprehensive genome-resolved metagenomics of nearly 1600 species originating from multiple anaerobic digesters. Biotechnol Biofuels 13(1):25. https://doi.org/10.1186/s13068-020-01679-y
6. Chu Y, Tu T, Penttinen L, Xue X, Wang X, Yi Z, Gong L, Rouvinen J, Luo H, Hakulinen N, Yao B, Su X (2017) Insights into the roles of non-catalytic residues in the active site of a GH10 xylanase with activity on cellulose. J Biol Chem 292(47):19315–19327. https://doi.org/10.1074/jbc.M117.807768
7. Colombo LT, De Oliveira MNV, Carneiro DG, De Souza RA, Alvim MCT, Dos Santos JC, Da Silva CC, Vidigal PMP, Da Silveira WB, Passos FML (2016) Applying functional metagenomics to search for novel lignocellulosic enzymes in a microbial consortium derived from a thermophilic composting phase of sugarcane bagasse and cow manure. Antonie Van Leeuwenhoek 109(9):1217–1233. https://doi.org/10.1007/s10482-016-0723-4
8. Colpa DI, Fraaije MW, Van Bloois E (2014) DyP-type peroxidases: a promising and versatile class of enzymes. J Ind Microbiol Biotechnol 41(1):1–7. https://doi.org/10.1007/s10295-013-1371-6
9. Dougherty MJ, D'haeseleer P, Hazen TC, Simmons BA, Adams PD, Hadi MZ (2012) Glycoside Hydrolases from a targeted Compost Metagenome, activity-screening and functional characterization. BMC Biotechnol 12(1):38. https://doi.org/10.1186/1472-6750-12-38
10. Ellilä S, Bromann P, Nyyssönen M, Itävaara M, Koivula A, Paulin L, Kruus K (2019) Cloning of novel bacterial xylanases from lignocellulose-enriched compost metagenomic libraries. AMB Express 9(1):124. https://doi.org/10.1186/s13568-019-0847-9
11. Ferrer M, Ghazi A, Beloqui A, Vieites JM, López-Cortés N, Marín-Navarro J, Nechitaylo TY, Guazzaroni M-E, Polaina J, Waliczek A, Chernikova TN, Reva ON, Golyshina OV, Golyshin PN (2012) Functional Metagenomics Unveils a Multifunctional Glycosyl Hydrolase from the Family 43 Catalysing the Breakdown of Plant Polymers in the Calf Rumen. PLoS ONE 7(6):e38134. https://doi.org/10.1371/journal.pone.0038134

12. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6(5):343–345. https://doi.org/10.1038/nmeth.1318
13. Gladden JM, Allgaier M, Miller CS, Hazen TC, VanderGheynst JS, Hugenholtz P, Simmons BA, Singer SW (2011) Glycoside Hydrolase Activities of Thermophilic Bacterial Consortia Adapted to Switchgrass. Appl Environ Microbiol 77(16):5804–5812. https://doi.org/10.1128/AEM.00032-11
14. Groves WE, Davis FC, Sells BH (1968) Spectrophotometric determination of microgram quantities of protein without nucleic acid interference. Anal Biochem 22(2):195–210. https://doi.org/10.1016/0003-2697(68)90307-2
15. Hogg BN, Schnepel C, Finnigan JD, Charnock SJ, Hayes MA, Turner NJ (2024) The Impact of Metagenomics on Biocatalysis. Angew Chem Int Ed 63(21):e202402316. https://doi.org/10.1002/anie.202402316
16. Jensen MS, Fredriksen L, MacKenzie AK, Pope PB, Leiros I, Chylenski P, Williamson AK, Christopeit T, Østby H, Vaaje-Kolstad G, Eijsink VGH (2018) Discovery and characterization of a thermostable two-domain GH6 endoglucanase from a compost metagenome. PLoS ONE 13(5):e0197862. https://doi.org/10.1371/journal.pone.0197862
17. Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, Kuhn K, Yuan J, Polevikov E, Smith TPL, Pevzner PA (2020) metaFlye: scalable long-read metagenome assembly using repeat graphs. Nat Methods 17(11):1103–1110. https://doi.org/10.1038/s41592-020-00971-x
18. Kolmogorov M, Yuan J, Lin Y, Pevzner PA (2019) Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37(5):540–546. https://doi.org/10.1038/s41587-019-0072-8
19. Lu M, Dukunde A, Daniel R (2019) Biochemical profiles of two thermostable and organic solvent–tolerant esterases derived from a compost metagenome. Appl Microbiol Biotechnol 103(8):3421–3437. https://doi.org/10.1007/s00253-019-09695-1
20. Menzel P, Ng KL, Krogh A (2016) Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun 7(1):11257. https://doi.org/10.1038/ncomms11257
21. Miller GL (1959) Use of dinitrosalicylic acid reagent for determination of reducing sugar. Anal Chem 31:426–428
22. Moss EL, Maghini DG, Bhatt AS (2020) Complete, closed bacterial genomes from microbiomes using nanopore sequencing. Nat Biotechnol 38(6):701–707. https://doi.org/10.1038/s41587-020-0422-6
23. Ndata K, Nevondo W, Cekuse B, Van Zyl LJ, Trindade M (2021) Characterization of a highly xylose tolerant β-xylosidase isolated from high temperature horse manure compost. BMC Biotechnol 21(1):61. https://doi.org/10.1186/s12896-021-00722-6
24. Nyathi M, Dhlamini Z, Ncube T (2024) Cloning Cellulase Genes from Victoria Falls Rainforest Decaying Logs Metagenome. Pol J Microbiol 73(3):343–348. https://doi.org/10.33073/pjm-2024-029
25. Pollegioni L, Tonin F, Rosini E (2015) Lignin-degrading enzymes. FEBS J 282(7):1190–1213. https://doi.org/10.1111/febs.13224
26. Reichart NJ, Bowers RM, Woyke T, Hatzenpichler R (2021) High Potential for Biomass-Degrading Enzymes Revealed by Hot Spring Metagenomics. Front Microbiol 12:668238.

https://doi.org/10.3389/fmicb.2021.668238

27. Rhee J-K, Ahn D-G, Kim Y-G, Oh J-W (2005) New Thermophilic and Thermostable Esterase with Sequence Similarity to the Hormone-Sensitive Lipase Family, Cloned from a Metagenomic Library. Appl Environ Microbiol 71(2):817–825. https://doi.org/10.1128/AEM.71.2.817-825.2005
28. Rhoads A, Au KF (2015) PacBio Sequencing and its Applications. Genomics Proteom Bioinf 13(5):278–289. https://doi.org/10.1016/j.gpb.2015.08.002
29. Robinson SL, Piel J, Sunagawa S (2021) A roadmap for metagenomic enzyme discovery. Nat Prod Rep 38(11):1994–2023. https://doi.org/10.1039/D1NP00006C
30. Sae-Lee R, Boonmee A (2014) Newly derived GH43 gene from compost metagenome showing dual xylanase and cellulase activities. Folia Microbiol (Praha) 59(5):409–417. https://doi.org/10.1007/s12223-014-0313-7
31. Saha BC (2001) Xylanase from a newly isolated Fusarium verticillioides capable of utilizing corn fiber xylan. Appl Microbiol Biotechnol 56(5–6):762–766. https://doi.org/10.1007/s002530100716
32. Santos A, Mendes S, Brissos V, Martins LO (2014) New dye-decolorizing peroxidases from Bacillus subtilis and Pseudomonas putida MET94: towards biotechnological applications. Appl Microbiol Biotechnol 98(5):2053–2065. https://doi.org/10.1007/s00253-013-5041-4
33. Santos-Pereira C, Sousa J, Costa ÂMA, Santos AO, Rito T, Soares P, Franco-Duarte R, Silvério SC, Rodrigues LR (2023) Functional and sequence-based metagenomics to uncover carbohydrate-degrading enzymes from composting samples. Appl Microbiol Biotechnol 107(17):5379–5401. https://doi.org/10.1007/s00253-023-12627-9
34. Sato M, Suda M, Okuma J, Kato T, Hirose Y, Nishimura A, Kondo Y, Shibata D (2017) Isolation of highly thermostable β-xylosidases from a hot spring soil microbial community using a metagenomic approach. DNA Res 24(6):649–656. https://doi.org/10.1093/dnares/dsx032
35. Schäfers C, Blank S, Wiebusch S, Elleuche S, Antranikian G (2017) Complete genome sequence of Thermus brockianus GE-1 reveals key enzymes of xylan/xylose metabolism. Stand Genomic Sci 12(1):22. https://doi.org/10.1186/s40793-017-0225-7
36. Seemann T (2014) Prokka: rapid prokaryotic genome annotation. Bioinformatics 30(14):2068–2069. https://doi.org/10.1093/bioinformatics/btu153
37. Sereika M, Kirkegaard RH, Karst SM, Michaelsen TY, Sørensen EA, Wollenberg RD, Albertsen M (2022) Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. Nat Methods 19(7):823–826. https://doi.org/10.1038/s41592-022-01539-7
38. Silva D, Sousa AC, Robalo MP, Martins LO (2023) A wide array of lignin-related phenolics are oxidized by an evolved bacterial dye-decolourising peroxidase. New Biotechnol 77:176–184. https://doi.org/10.1016/j.nbt.2022.12.003
39. Soghomonyan T, Hambardzumyan A, Mkhitaryan A, Khoyetsyan L, Paronyan M, Izmailyan M, Kinosyan M, Bagiyan V, Ghochikyan V, Panosyan H, Paloyan A (2024) Obtaining and Characterizing Thermostable α-Amylases Secreted by Bacillus subtilis, Originating from Bacillus amyloliquefaciens and Bacillus subtilis. Fermentation 10(11):547. https://doi.org/10.3390/fermentation10110547

40. Spilliaert R, Hreggvidsson GO, Kristjansson JK, Eggertsson G, Palsdottir A (1994) Cloning and Sequencing of a *Rhodothermus marinus* Gene, *bglA*, Coding for a Thermostable β-Glucanase and its Expression in *Escherichia coli*. Eur J Biochem 224(3):923–930. https://doi.org/10.1111/j.1432-1033.1994.00923.x
41. Sugano Y, Yoshida T (2021) DyP-Type Peroxidases: Recent Advances and Perspectives. Int J Mol Sci 22(11):5556. https://doi.org/10.3390/ijms22115556
42. Tirawongsaroj P, Sriprang R, Harnpicharnchai P, Thongaram T, Champreda V, Tanapongpipat S, Pootanakit K, Eurwilaichitr L (2008) Novel thermophilic and thermostable lipolytic enzymes from a Thailand hot spring metagenomic library. J Biotechnol 133(1):42–49. https://doi.org/10.1016/j.jbiotec.2007.08.046
43. Uchiyama T, Miyazaki K, Yaoi K (2013) Characterization of a Novel β-Glucosidase from a Compost Microbial Metagenome with Strong Transglycosylation Activity. J Biol Chem 288(25):18325–18334. https://doi.org/10.1074/jbc.M113.471342
44. Vishan I, Sivaprakasam S, Kalamdhad A (2017) Isolation and identification of bacteria from rotary drum compost of water hyacinth. Int J Recycl Org Waste Agric 6(3):245–253. https://doi.org/10.1007/s40093-017-0172-8
45. Wu Y-W, Simmons BA, Singer SW (2016) MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32(4):605–607. https://doi.org/10.1093/bioinformatics/btv638
46. Xia Y, Ju F, Fang HHP, Zhang T (2013) Mining of Novel Thermo-Stable Cellulolytic Genes from a Thermophilic Cellulose-Degrading Consortium by Metagenomics. PLoS ONE 8(1):e53779. https://doi.org/10.1371/journal.pone.0053779
47. Zhang Y, Schmid YRF, Luginbühl S, Wang Q, Dittrich PS, Walde P (2017) Spectrophotometric Quantification of Peroxidase with *p* -Phenylene-diamine for Analyzing Peroxidase-Encapsulating Lipid Vesicles. Anal Chem 89(10):5484–5493. https://doi.org/10.1021/acs.analchem.7b00423

Figures

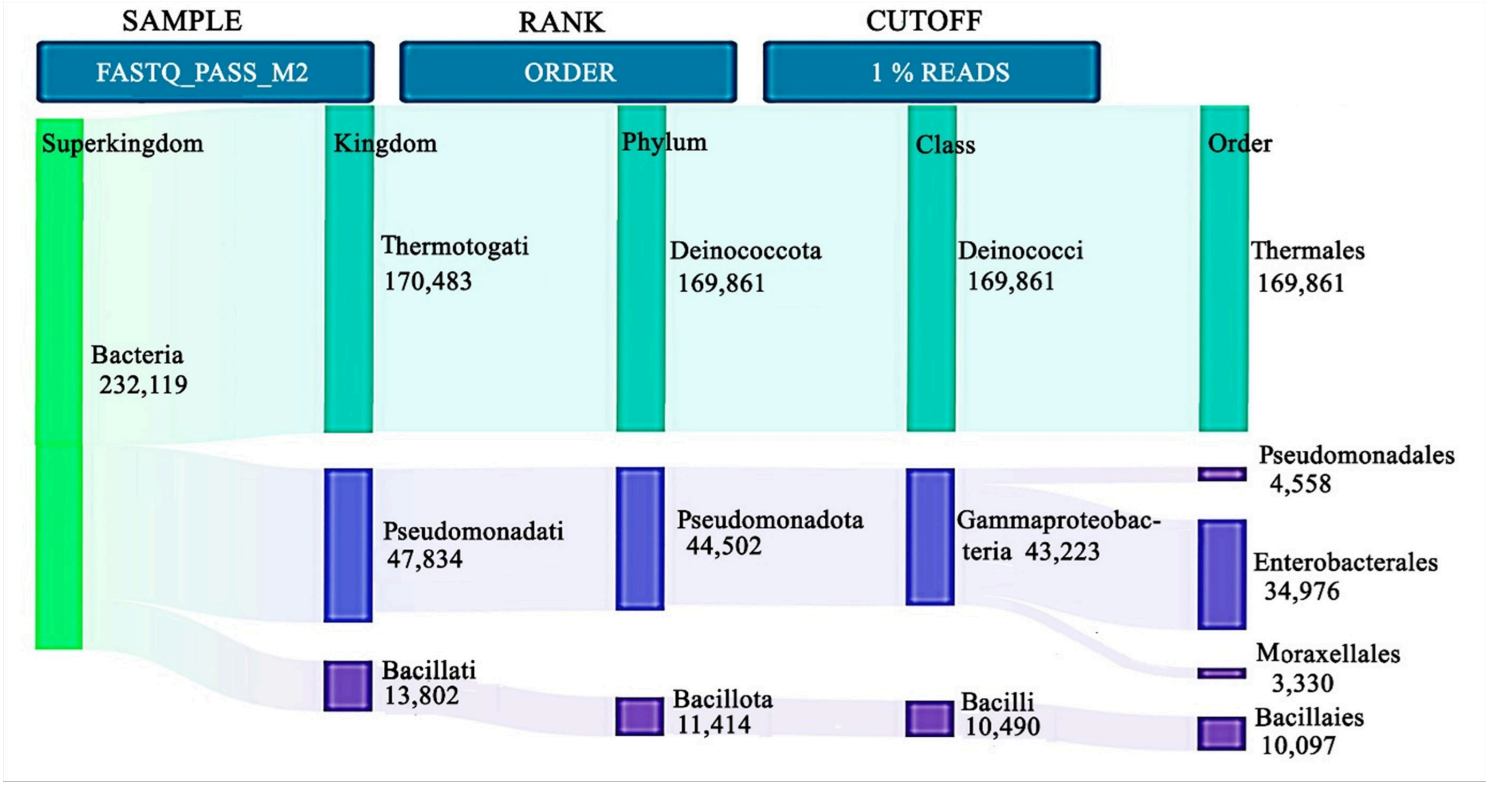


Figure 1

Schematic representation of the bacterial superkingdom composition based on 317,393 high-quality reads generated by the EPI2ME Desktop metagenomics workflow.

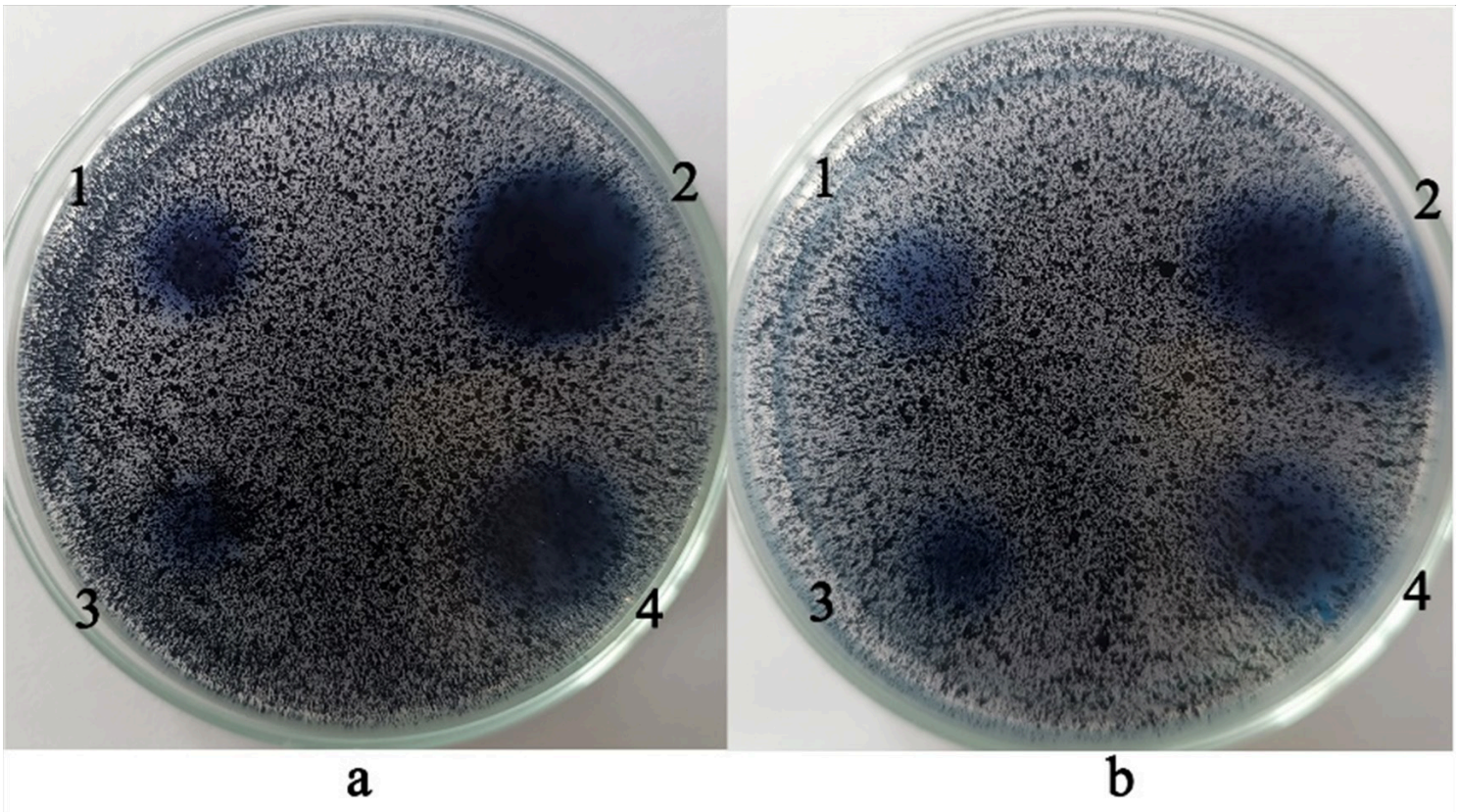


Figure 2

Estimation of extracellular xylanase activity of metagenome-derived gene products NGJGEFEI_14874 (14) and NGJGEFEI_34975 (34). Reaction conditions: as described in Materials and Methods. (a) 14: 1 – 5 µL of CL from strain 14; 2 – 20 µL of CL from strain 14; 3 – 5 µL of CL from strain RIK; 4 – 20 µL of CL from strain RIK. (b) 34: 1 – 5 µL of CL from strain 34; 2 – 20 µL of CL from strain 34; 3 – 5 µL of CL from strain RIK; 4 – 20 µL of CL from strain RIK.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformationinfo.docx
- Supplementarymaterials.docx