ISSN: 2621-4709 E-ISSN: 2621-2528
Journal of Applied Agricultural Science and Technology
Vol. 9 No. 2 (2025): 286-305

Metagenomic Bioprospecting for Lignocellulosic Enzymes from Bacterial Communities of Humus Obtained from Natural and Man-Made Forests in Tomohon, North Sulawesi, Indonesia

Feky Recky Mantiri [a,*], Carla Felly Kairupan [b], Sri Sudewi [c], Vic Axel Daniel Mantiri [d]

[a] Department of Biology, Faculty of Mathematics and Natural Sciences, Universitas Sam Ratulangi, Manado, Indonesia
[b] Study Program of Medical Doctor Education, Faculty of Medicine, Universitas Sam Ratulangi, Manado, Indonesia
[c] Department of Pharmacy, Faculty of Mathematics and Natural Sciences, Universitas Sam Ratulangi, Manado, Indonesia
[d] Study Program of Medical Doctor Profession, Faculty of Medicine, Universitas Sam Ratulangi, Manado, Indonesia

Abstract. *Lignocellulosic biomass degradation is crucial for various industrial applications. Traditional enzyme discovery methods, limited by culturing constraints, fail to capture the vast enzymatic potential of uncultured microorganisms. Metagenomic bioprospecting provides a culture-independent avenue to explore this untapped genetic diversity. This research characterizes the microbial communities and their functional capabilities in a natural forest (Mahawu Mountain Forest, MMF) and a man-made forest (Tomohon City Forest, TCF) located in North Sulawesi, Indonesia, aiming to assess the influence of forest type on microbial ecological dynamics and lignocellulose degradation mechanisms. Comparative soil analysis revealed MMF had slightly alkaline pH (7.1), cooler temperature (21°C), and dark grayish-brown Andosol, while TCF exhibited a neutral pH (6.9), warmer temperature (23°C), and brown Andosol. High-throughput 16S rRNA sequencing demonstrated that TCF harbors greater bacterial richness (125 vs. 91 observed OTUs) and diversity (Shannon index 4.44 vs. 4.11), likely influenced by anthropogenic activities. Taxonomic profiling showed that Proteobacteria dominate both sites (MMF: 42.37%; TCF: 56.08%), with Actinobacteria significantly more abundant in MMF (34.08% vs. 5.84%). Functional prediction via PICRUSt analysis highlighted TCF's enrichment in stress-responsive genes and ABC transporters, whereas MMF exhibited elevated lipid metabolism and specialized lignin-degradation pathways (e.g., 3-hydroxyphenylacetate degradation). These findings suggest that TCF's heterogeneous environment supports microbial versatility, while MMF's stable conditions promote specialization in decomposition. Both forests represent promising reservoirs for lignocellulolytic enzyme discovery, with implications for sustainable biotechnological applications. This study underscores the importance of forest management in shaping soil microbial communities and highlights the value of preserving natural ecosystems for future bioresource exploration.*
Keywords: 16S rRNA sequencing; bioprospecting; metagenomics; soil microbiome.

Type of the Paper: Regular Article.

1. Introduction

The global transition toward renewable energy and sustainable bioproducts has made lignocellulosic biomass a primary focus in bioeconomy research [1]. Lignocellulose is a complex

https://doi.org/10.55043/jaast.v9i2.421
Received March 12, 2025; Received in revised form May 2, 2025; Accepted May 19, 2025; Published May 29, 2025
* First corresponding author
Email: fmantiri@unsrat.ac.id
© 2024 The Authors. Published by Green Engineering Society on Journal of Applied Agricultural Science and Technology
This is an open access article under the CC BY-SA 4.0 license https://creativecommons.org/licenses/by-sa/4.0

and abundant biopolymer composed of cellulose, hemicellulose, and lignin, which together form the structural framework of plant cell walls. The abundant lignocellulose resource remains difficult to process efficiently due to its natural resistance, requiring enzyme-based hydrolysis to obtain fermentable sugars needed for biofuel production, bioplastic manufacturing, and other valuable products [2–4]. Research into new lignocellulolytic enzymes that offer improved performance metrics such as activity, stability, and specificity remains essential for enhancing the efficiency and cost-effectiveness of biomass conversion processes [5–7].

Soil ecosystems, particularly humus-rich layers, are hotspots of microbial diversity and enzymatic activity [8]. Microbial communities vital to nutrient cycling and organic matter degradation are concentrated in humus, which is produced by the breakdown of organic materials. Numerous lignocellulosic enzymes found in these microbial consortia have application in various biotechnological fields, including waste management and biofuel production. Bacteria are recognized for their metabolic adaptability and their capacity to synthesize a diverse array of lignocellulolytic enzymes, encompassing cellulases, xylanases, and ligninases [9–11].

The identification of genes encoding lignocellulolytic enzymes from various habitats, such as soil, compost, and herbivore gut microbiomes, has been facilitated by recent developments in bioinformatics and high-throughput sequencing [2,12,13]. Metagenomics, in particular, has become a valuable tool for investigating the genetic potential of microbial populations. By combining metagenomic sequencing with functional annotation, researchers can find novel enzymes with special qualities, including strong thermal stability or activity under extreme pH settings, which are essential for industrial applications [14].

Tropical areas, including North Sulawesi, Indonesia, are valuable sources for bioprospecting new enzymes due to their distinctive environmental conditions and rich biodiversity [14–16]. The city of Tomohon, located in North Sulawesi, contains unique natural and man-made forests with distinct ecological dynamics. This region provides a diverse range of microbial life that can support biotechnological discoveries and potentially lead to the identification of enzymes capable of advancing a number of industrial processes, including the production of medications and biofuels. Thus far, soil microbial bioprospecting research in this area remains limited. In addition to improving our understanding of microbial diversity, investigating these varied habitats creates new paths for environmentally friendly sustainable enzyme production and utilization. Therefore, this study aims to characterize bacterial diversity in humus from natural and man-made forests in Tomohon, Indonesia, and to identify lignocellulolytic enzymes through metagenomic sequencing.

2. Materials and Methods

2.1. Sampling

Soil samples were collected in September 2023 from Mount Mahawu Protected Forest (MMF) (1.3507621N, 124.8655716E) and Tomohon City Forest (TCF) (1.3340270N, 124.8491224E) in Talete, Tomohon City, North Sulawesi, Indonesia (Fig. 1).

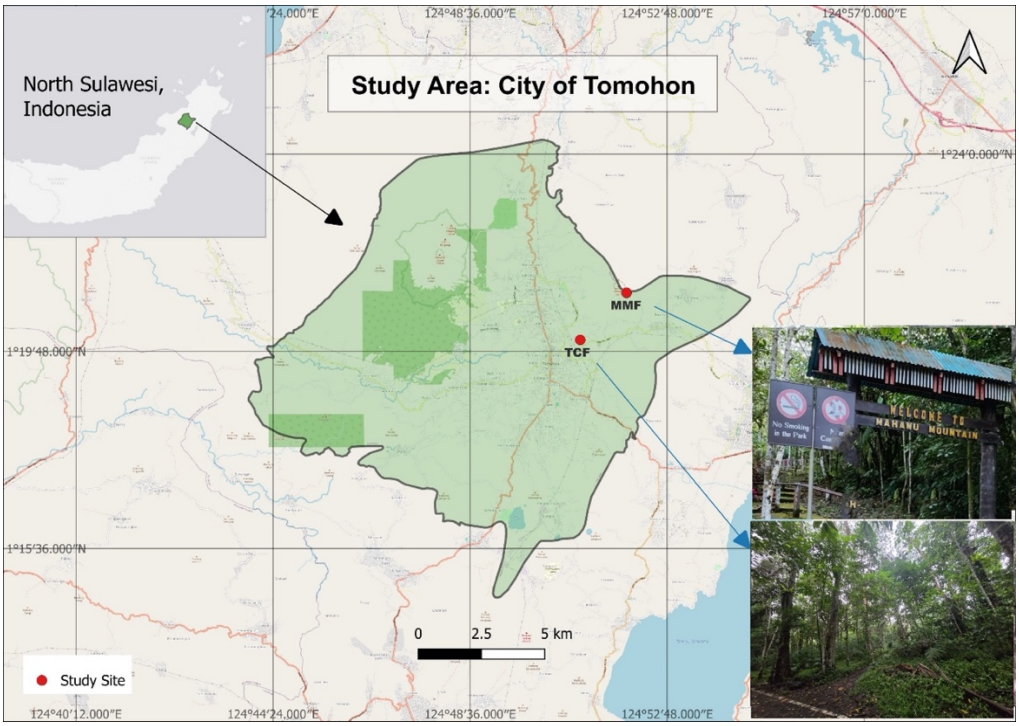

Fig. 1. Map of the study area. Mount Mahawu Forest (MMF) and Tomohon City Forest (TCF) are located in the City of Tomohon, North Sulawesi Indonesia.

Random sampling method was employed to ensure that all areas of each forest type were represented, effectively capturing the spatial variability of the field [17]. Composite samples were collected from the top 0-20 cm depth using sterilized spatulas to prevent cross-contamination. At each site, surface vegetation and the top 2 cm of soil were removed to avoid plant material contamination. Six composite samples, consisting of multiple pooled subsamples from within a defined area, were homogenized, and sieved through a sterile 2 mm mesh to remove debris and stones. Samples were immediately transferred to sterile 50-mL Falcon tubes. To minimize microbial activity and preserve the original community structure, samples were stored on ice during transport and subsequently frozen at -20°C for long-term storage until further processing.

Soil physical parameters—altitude, pH, temperature, and relative humidity—were measured directly in the field using standard equipment and procedures, while chemical parameters were analyzed in the Laboratory of Soil, Faculty of Agriculture, Sam Ratulangi University.

2.2. Genomic DNA Extraction, Library Preparation, and Sequencing

Humus samples were analyzed at PT. Genetika Science, Indonesia, employing NGS-16 S Sequencing. Briefly, genomic DNA was extracted from 0.25 g of each composite soil sample using

the ZymoBIOMICS™ DNA Miniprep Kit (Zymo Research, USA), following the manufacturer's instructions. The V3-V4 regions of the 16S rRNA gene was amplified using primers 341F (CCTACGGGNGGCWGCAG) and 805R (GACTACHVGGGTATCTAATCC) with Illumina adapters [18]. PCR conditions included an initial denaturation at 95°C for 3 minutes, followed by 25 cycles of 95°C for 30 seconds, 55°C for 30 seconds, and 72°C for 30 seconds, with a final extension at 72°C for 5 minutes. Amplicon libraries were prepared using the NEBNext® Ultra™ DNA Library Prep Kit (Illumina, USA) and sequenced on the Illumina HiSeq 2500 platform (2 × 250 bp paired-end reads) to generate paired-end raw reads.

2.3. Analysis of Data

Adapters, primers, and low-quality sequences (average score < 20, read length < 100 bp) were removed using Trimmomatic v0.39 and Cutadapt v2.10 [19]. Reads were clustered and dereplicated into amplicon sequence variants (ASVs) using the DADA2 plugin [20] within the QIIME2 pipeline (version 2020.8). Taxonomic profiles were analyzed using QIIME2 against the SILVA 132 database (http://www.arb-silva.de/). Alpha and beta diversity metrics were calculated in QIIME2, with statistical comparisons performed using PERMANOVA and paired t-tests in R v4.2.3 (https://www.R-project.org/). Community analyses included alpha diversity (Shannon-Weiner index) and beta diversity assessment through Principal Coordinate Analysis (PCoA), performed using R and Minitab 19. Functional profiles were deduced using PICRUSt2 v2.3.0-b in R v4.2.3.

3. Results and Discussion

3.1. Soil type, Physical, and Chemical Properties of Soil in MMF and TCF

Comparative analysis of soil properties between the natural forest (MMF) and the man-made forest (TCF) exhibit distinct differences in the physical and chemical properties of their soils. Situated at an altitude of 1262 meters above sea level (asl), the MMF features Andosol soil, characterized by a dark grayish-brown color. The soil pH is slightly alkaline at 7.1, with a temperature of 21°C and a humidity level of 45%. In contrast, the TCF, located at a lower altitude of 899 meters asl, also has Andosol soil but with a distinct brown color. The pH of the TCF soil is slightly lower at 6.9, indicating a more neutral to slightly acidic environment. The temperature in the TCF is warmer at 23°C, and the humidity is slightly higher at 46%. These differences in soil properties may influence the types of vegetation and microbial communities present in each forest. The cooler and more alkaline conditions of the MMF soil could support different plant species compared to the warmer and slightly acidic soil of the TCF. Additionally, the variations in humidity and temperature could affect soil moisture retention and evaporation rates [21], further shaping the ecological dynamics of these two forest ecosystems.

3.2. Sequencing Quality

The dataset in the current report provides information on bacterial diversity of humus from two distinct forest ecosystems in North Sulawesi, Indonesia, namely MMF, which represents a natural forest, and TCF, which represents a man-made forest. Initial sequencing employing high-throughput sequencing of the 16S rRNA gene sequencing yielded 488,370 and 438,090 paired-end reads for TCF and MMF, respectively (Table 1).

Table 1. Sequencing Statistics of Metagenomes from MMF and TCF.

Sample Name	Raw PE (#)	Raw Tags (#)	Clean Tags (#)	Effective Tags (#)	Base(nt)	AvgLen (nt)	Q20	Q30	GC%	Effective %
Mount Mahawu Forest (MMF)	438,090	219,045	218,819	218,819	62,120,281	283.90	89.10	77.30	57.30	49.94%
Tomohon City Forest (TCF)	488,370	244,185	244,066	244,066	69,295,204	283.90	89.20	77.50	57.50	49.97%

Quality metrics demonstrated exceptional sequencing fidelity across both sampling sites. Base quality scores exceeded standard thresholds, with Q20 values ranging from 89.1 to 89.2% and Q30 values ranging from 77.3 to 77.5%. The remarkably consistent GC content (57.3-57.5%) between sites indicates minimal bias in DNA extraction and amplification procedures. Read length uniformity was maintained at 283.9 nucleotides across both sites, confirming successful targeted amplification of the 16S rRNA gene region. Post-quality filtering retention rates were notably high, with 99.90% and 99.95% of sequences retained for MMF and TCF, respectively. The effective percentage (~50%) for both sites represents the proportion of high-quality sequences retained after paired-end merging, aligning with expected values for 16S amplicon sequencing protocols. The high-quality dataset, reflected by low technical variation between sites, provides a robust foundation for future comparative microbial community analyses and diversity.

3.3. Diversity and composition of bacterial communities in MMF and TCF.

3.3.1. Bacterial Diversity and Richness

Rarefaction analysis (Fig. 2) illustrates differences in operational taxonomic unit (OTU) abundance between MMF and TCF. The curve for TCF approached a higher asymptote, indicating greater OTU richness (125 observed OTUs) compared to MMF (91 OTUs) (Table 2). This suggests that the man-made forest harbors a more diverse bacterial community, potentially due to anthropogenic influences such as land management or nutrient enrichment. Diversity indices further supported this observation, with TCF exhibiting a higher Shannon index (4.44) compared to MMF (4.11), reflecting greater evenness and richness. The Simpson index (0.983 for TCF vs. 0.979 for MMF) and inverse Simpson index (58.44 for TCF vs. 47.11 for MMF) similarly indicated lower taxonomic dominance and a more equitable distribution in TCF. The Venn diagram (Fig. 3)

revealed both shared and unique OTUs, with TCF displaying a higher proportion of exclusive OTUs, suggesting the presence of specialized taxa adapted to its altered ecological conditions. These differences highlight the impact of forest type on microbial diversity, with TCF's heterogeneity likely promoting novel niches [22].

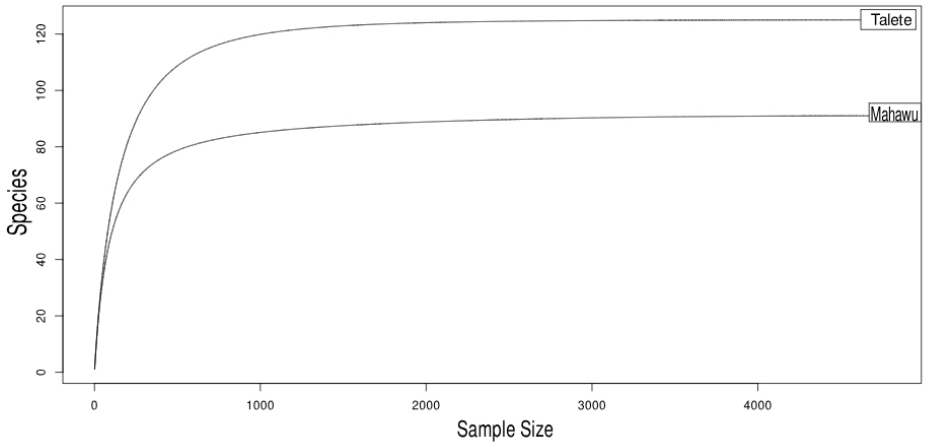

Fig. 2. Rarefaction Plot reflecting the variation in OTUs abundance for each location. Mahawu and Talete represent MMF and TCF, respectively.

Table 2. Microbiome alpha diversity metrics for MMF and TCF.

	Observed	Shannon	Simpson	InvSimpson
MMF	91	4.10965072843096	0.978775095903672	47.1144649446494
TCF	125	4.44140190183787	0.982888982072048	58.4418767025199

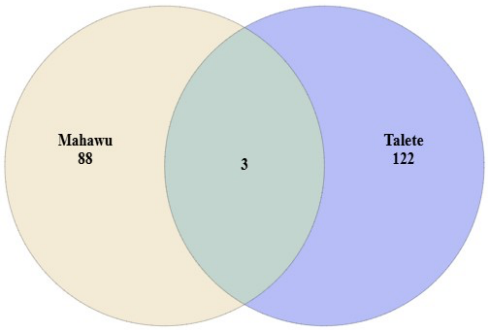

Fig. 3. Venn diagram showing the number of unique and shared OTUs in both locations. Mahawu and Talete represent MMF and TCF, respectively.

3.3.2. Taxonomic Composition Across Hierarchical Levels

Taxonomic profiling revealed both similarities and distinctions in bacterial composition between MMF and TCF across phylum, class, order, family, and genus levels, reflecting ecological and functional specialization.

At the phylum level (Fig. 4), Proteobacteria dominated both ecosystems, comprising 42.37% in MMF and 56.08% in TCF, reflecting their metabolic versatility in lignocellulose degradation [23]. Actinobacteriota was notably higher in MMF (34.08%) than in TCF (5.84%), suggesting a stronger role in natural forest decomposition. Planctomycetota accounted for 5.90% in MMF and 6.84% in TCF, while Acidobacteriota was more abundant in TCF (8.54%) than in MMF (3.79%). Verrucomicrobiota (4.38% in MMF, 3.54% in TCF), Firmicutes (3.64% in MMF, 3.98% in TCF),

and Myxococcota (3.22% in MMF, 2.47% in TCF) were present in both. However, Bacteroidota (7.63% in TCF vs. 2.62% in MMF), Gemmatimonadota (1.83% in TCF, absent in MMF), and Nitrospirota (3.24%) in TCF, absent in MMF) showed TCF-specific enrichment, indicating broader phylum-level diversity.

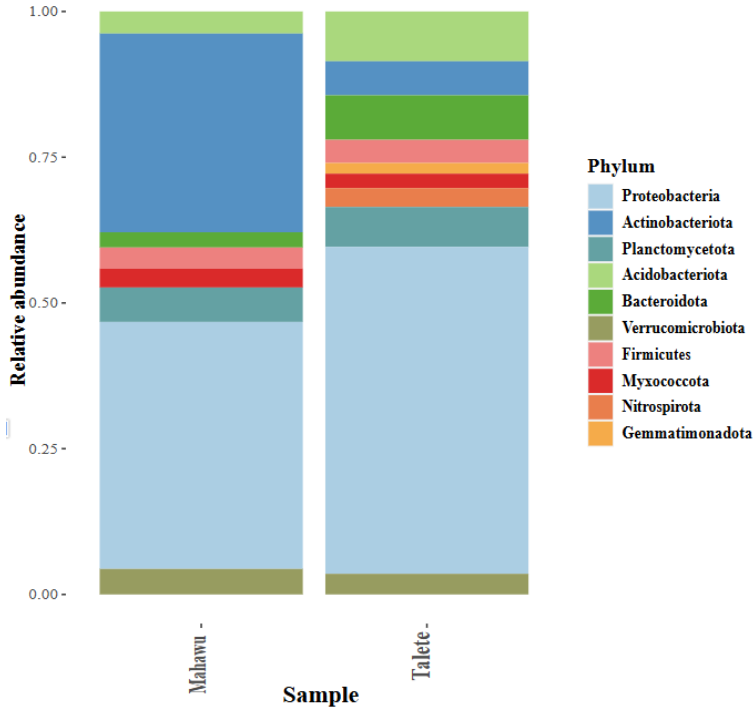

Fig. 4. Relative abundance of the ten dominant bacterial phyla identified from both locations. Mahawu and Talete represent MMF and TCF, respectively.

The dominance of Proteobacteria in both MMF and TCF aligns with their well-documented roles in lignocellulose degradation [24], as highlighted by studies on soil microbial communities. Research consistently shows that Proteobacteria are prevalent in environments rich in organic matter, where they play crucial roles in carbon cycling [25]. This phylum's metabolic versatility allows it to thrive in diverse environments and efficiently degrade complex organic matter. However, the significantly higher abundance of Actinobacteriota in MMF suggests a stronger role in natural forest decomposition. Actinobacteria are known for producing a wide array of hydrolytic enzymes, making them key contributors to the breakdown of recalcitrant plant polymers. In contrast, the markedly lower abundance of Actinobacteriota in TCF suggests differing decomposition pathways or nutrient cycling dynamics. This trend aligns with findings that Actinobacteria abundance can vary significantly with soil pH, moisture, and vegetation type [26].

The enrichment of Bacteroidota, Gemmatimonadota, and Nitrospirota specifically in TCF indicates a broader phylum diversity compared to MMF. Bacteroidota are often associated with polysaccharides degradation, while Gemmatimonadota and Nitrospirota contribute to nitrogen and sulfur cycling, respectively. This suggests that TCF may support a more complex nutrient cycling network, likely influenced by its higher moisture content and distinct vegetation. Studies have

shown that Bacteroidota are often more abundant in environments with high organic matter input and limited oxygen availability, consistent with TCF characteristics [27].

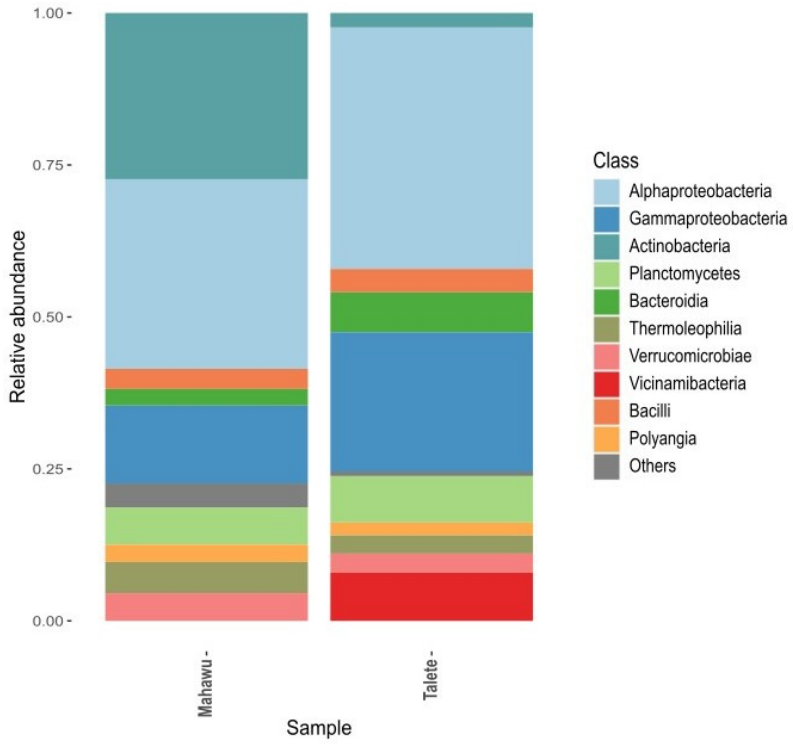

Fig. 5. Relative abundance of the ten dominant bacterial classes identified from both locations. Mahawu and Talete represent MMF and TCF, respectively.

Class-level analysis (Fig. 5) identified Alphaproteobacteria as dominant, comprising 31.21% in MMF and 39.71% in TCF, followed by Actinobacteria (27.34% in MMF; 2.40% in TCF), consistent with phylum trends. Gammaproteobacteria increased from 12.81% in MMF to 22.91% in TCF, reflecting nutrient-rich conditions in the latter. Planctomycetes (6.13% in MMF; 7.64% in TCF) and Bacteroidia (2.73% in MMF; 6.57% in TCF) were also notable. Bacilli (3.30% in MMF; 3.83% in TCF), Thermoleophilia (5.14% in MMF; 3.00% in TCF), and Verrucomicrobiae (4.55% in MMF, 3.16% in TCF) showed modest shifts. Polyangia (2.86% in MMF; 2.12% in TCF) remained consistent, while the presence of Vicinamibacteria (7.90% in TCF, absent in MMF) underscored TCF's distinct taxonomic profile.

Order-level data (Fig. 6) highlighted Rhizobiales as dominant, comprising 33.40% in MMF and 39.90% in TCF, followed by Frankiales (26.90% in MMF; 2.20% in TCF) and Burkholderiales (13.30% in MMF, 27.20% in TCF). Bacillales (3.90% in MMF; 4.00% in TCF) and Solirubrobacterales (5.00% in MMF, 3.60% in TCF) remained stable, while Pedosphaerales (5.40% in MMF; 1.40% in TCF), Gemmatales (3.70% in MMF; 2.80% in TCF), and Pirellulales (3.00% in MMF; 5.70% in TCF) showed variations. Microtrichales (3.50% in MMF; 1.30% in TCF) declined in TCF, wherease Vicinamibacterales (9.30% in TCF, absent in MMF) emerged as a TCF-specific order, suggesting ecological specialization.

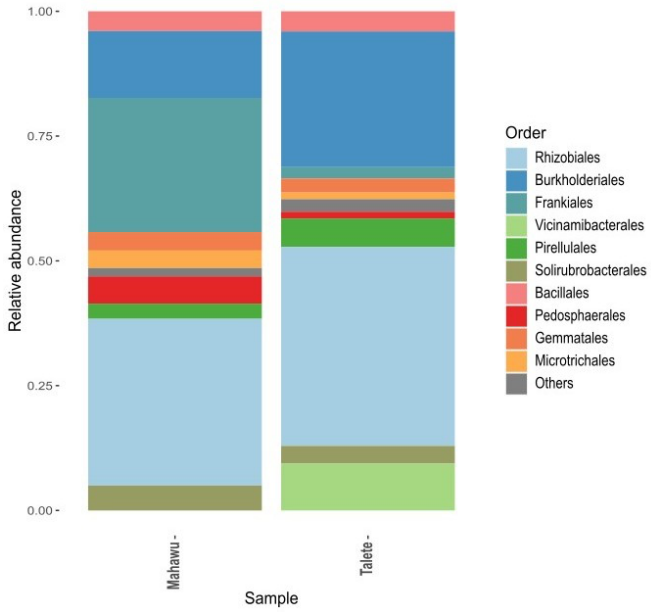

Fig. 6. Relative abundance of the ten dominant bacterial orders identified from both locations. Mahawu and Talete represent MMF and TCF, respectively.

At the class level, the dominance of Alphaproteobacteria in both ecosystems, followed by Actinobacteria in MMF and Gammaproteobacteria in TCF, further underscores the functional specializations observed at the phylum level. The increase in Gammaproteobacteria in TCF likely reflects its nutrient-rich conditions, as this class is known for its adaptability to diverse substrates [28]. The high abundance of Frankiales in MMF, compared to TCF, likely reflects the nitrogen-fixing capabilities of this order, which are essential in less disturbed natural forest ecosystems. However, the dominance of Rhizobiales in both forests supports the widespread role of nitrogen cycling at both sites. Rhizobiales are known for their symbiotic nitrogen-fixing abilities, particularly with legumes, which may be present in both forest types [29].

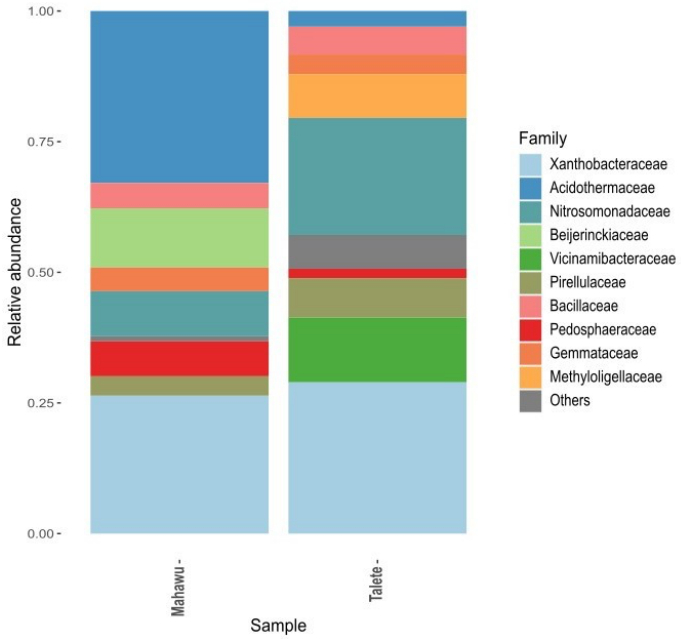

Fig. 7. Relative abundance of the ten dominant bacterial families identified from both locations. Mahawu and Talete represent MMF and TCF, respectively.

The emergence of Vicinamibacterales as a TCF-specific order highlights the unique ecological niche of this forest type. Although relatively understudied, its presence suggests a potential role in specialized decomposition or nutrient cycling processes unique to TCF [30].

At the family level (Fig. 7), Acidothermaceae dominated in MMF (32.91%) but declined to 2.99% in TCF, while Xanthobacteraceae was prominent in both (26.38% in MMF; 28.95% in TCF). Nitrosomonadaceae increased from 8.63% in MMF to 22.40% in TCF, and Bacillaceae remained stable (4.85% in MMF; 5.35% in TCF). Pedosphaeraceae (6.69% in MMF; 1.80% in TCF), Pirellulaceae (3.72% in MMF; 7.52% in TCF), and Gemmataceae (4.53% in MMF; 3.74% in TCF) showed shifts, while Beijerinckiaceae (11.32% in MMF, absent in TCF), Methyloligellaceae (8.38% in TCF, absent in MMF), and Vicinamibacteraceae (12.42% in TCF, absent in MMF) highlighted site-specific patterns.

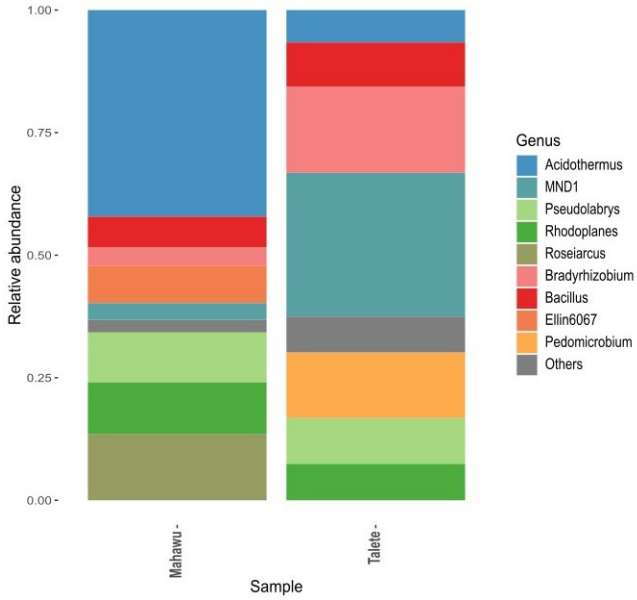

Fig. 8. Relative abundance of the ten dominant bacterial genera identified from both locations. Mahawu and Talete represent MMF and TCF, respectively.

Genus-level analysis (Fig. 8) identified *Acidothermus* as dominant in MMF (42.14%) but reduced in TCF (6.61%), reflecting Actinobacteria trends. *MND1* (2.93% in MMF; 29.31% in TCF) and *Bradyrhizobium* (3.77% in MMF; 17.51% in TCF) surged in TCF, while *Bacillus* (6.21% in MMF; 9.08% in TCF) and *Pseudolabrys* (10.18% in MMF; 9.33% in TCF) were notable. *Rhodoplanes* (10.60% in MMF; 7.43% in TCF) and *Roseiarcus* (13.45% in MMF, absent in TCF) favored MMF, whereas *Pedomicrobium* (13.38% in TCF, absent in MMF) and *Ellin6067* (7.66% in MMF, absent in TCF) were site-specific. *Others* (2.61% in MMF, 7.35% in TCF) indicated additional diversity.

The contrasting shift in Acidothermaceae abundance between MMF and TCF, with Acidothermus dominating MMF, suggests a strong influence of environmental factors on the distribution of this family. Acidothermus species are known for their thermophilic and cellulolytic

capabilities, making them promising targets for bioprospecting [31]. The surge of Nitrosomonadaceae in TCF, particularly the genus MND1, indicates a significant role in nitrogen oxidation, likely driven by higher moisture and nutrient availability. The abundance of Nitrosomonadaceae is often associated with ammonia oxidation, a key step in the nitrogen cycle [32].

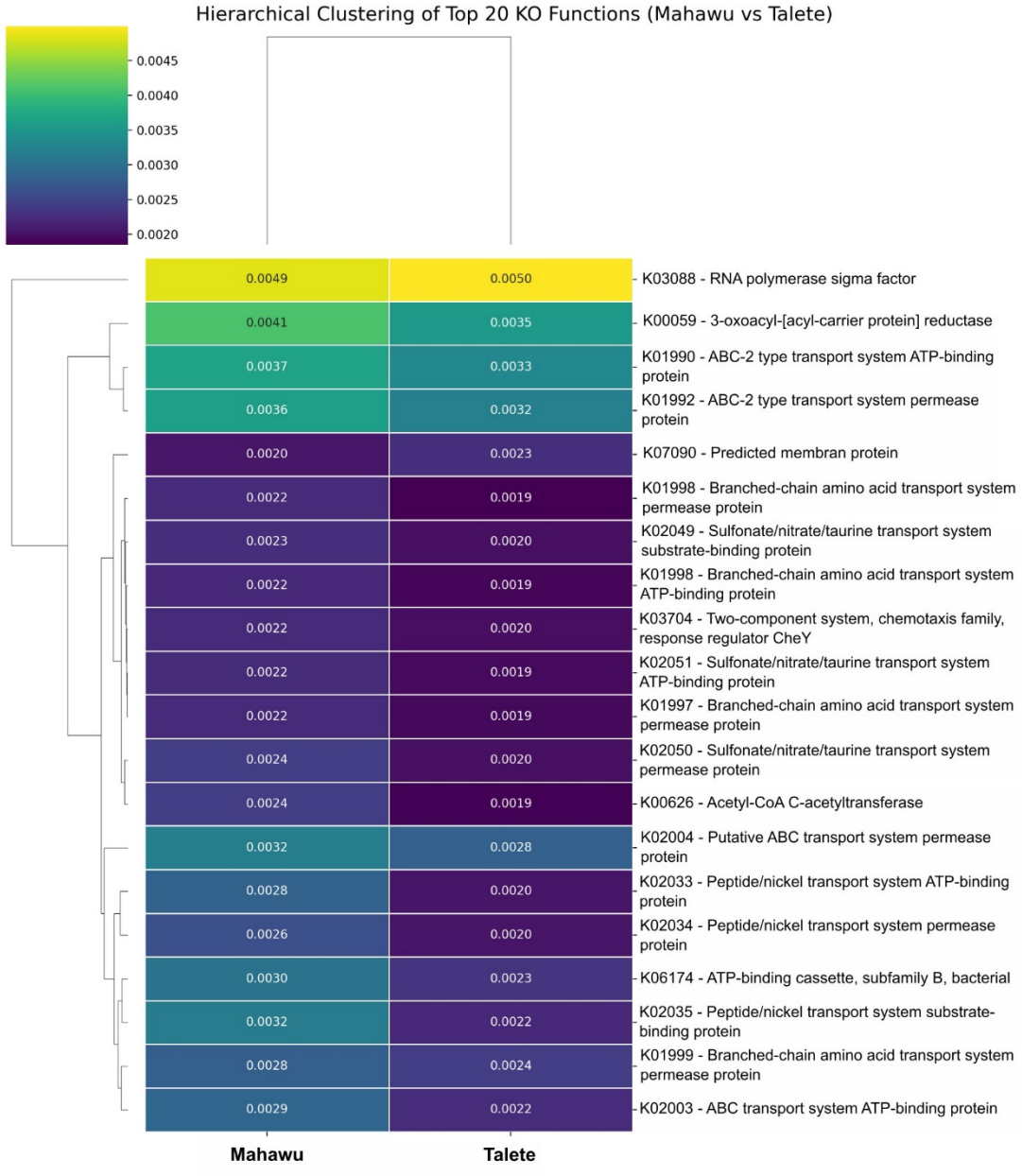

Fig. 9. KEGG-based PICRUSt analysis showing the relative abundance of the predicted functional gene content of microbial communities in Mahawu (MMF) and Talete (TCF). Pathway abundance values (yellow, green, blue, and purple) represent the number of genes from highest to lowest, respectively, and are normalized to the total number of genes present in a given pathway from each sample.

The increased abundance of *Bradyrhizobium* in TCF suggests enhanced nitrogen fixation in this environment. The presence of *Bacillus* in both ecosystems, with a slight increase in TCF, indicates its ubiquitous role in decomposition and nutrient cycling. The site-specific presence of

Pedomicrobium and Ellin6067 in TCF and MMF, respectively, highlights the unique microbial communities associated with each forest type. Studies have shown that *Pedomicrobium* is often found in acidic soils and is involved in manganese oxidation, which could explain its prevalence in TCF [33].

3.4. Functional Gene Prediction

To predict the functional potential of each microbial community, a KEGG-based PICRUSt analysis was conducted. In this analysis, the predicted functional gene composition of the microbial community is categorized into KEGG orthologs (KOs), which are subsequently mapped to KEGG pathways. The relative abundance of each KEGG pathway is determined based on the anticipated KO composition of the community. The findings from the KEGG-based PICRUSt, showing the relative abundance of functional gene composition within each forest type, are illustrated in Fig. 9.

The hierarchical clustering analysis of the top 20 KEGG Orthology (KO) functions revealed distinct functional profiles in the microbial communities of Mahawu (natural forest, MMF) and Talete (man-made forest, TCF), indicating potential adaptations to their respective ecological niches. ABC transporter systems dominated the functional landscape in both communities, with KO functions such as K01990 (ABC-2 type transport ATP-binding protein) and K01992 (ABC-2 type permease protein) exhibiting high abundances. These transporters are critical for the uptake of oligosaccharides, amino acids, and other nutrients derived from lignocellulose degradation [34]. Notably, the Mahawu's community exhibited slightly higher abundances of several transporters (e.g., K01990: 0.0037 vs. 0.0033 in Talete), suggesting enhanced substrate acquisition strategies potentially correlated with the natural forest's diverse organic matter inputs. Conversely, Talete's microbial profile showed a higher relative abundance of K03088 (RNA polymerase sigma factor), a regulator of stress-responsive genes [35] (0.0050 vs. 0.0049 in Mahawu), which may reflect adaptive mechanisms for coping with environmental fluctuations in the managed ecosystem.

Lipid metabolism-related enzymes, such as K00059 (3-oxoacyl-[acyl-carrier protein] reductase) and K00626 (acetyl-CoA C-acetyltransferase), were more prominent in Mahawu. These enzymes likely support membrane synthesis or redox balancing during lignocellulose degradation—processes requiring substantial energy expenditure [36]. Evidence of chemotaxis and environmental sensing was also present, as indicated by the detection of K03704 (CheY response regulator) in both communities, underscoring the importance of motility for locating and colonizing nutrient-rich substrates within humus [37]. A notable divergence was observed in membrane-associated functions: Talete's community exhibited elevated abundance of K07090 (predicted membrane protein) [38], potentially reflecting structural adaptations to secondary metabolites or altered substrate availability in the man-made forest.

The prevalence of ABC transporters and redox-related enzymes implies a metabolic focus on nutrient scavenging and energy conservation [39], indirectly supporting lignocellulose decomposition. Although canonical lignocellulolytic enzymes (e.g., cellulases, ligninases) were absent from the top KO functions, the observed transport systems suggest active uptake of hydrolyzed oligomers—a critical step in downstream degradation. The functional divergence between sites may reflect ecological strategies shaped by their environments: Mahawu's natural forest microbiome appears optimized for substrate diversity, whereas Talete's community prioritizes stress tolerance and regulatory flexibility. However, the absence of Carbohydrate-Active Enzymes (CAZymes) among the top KO functions highlights a limitation of the current analysis. More comprehensive metagenomic sequencing or targeted functional assays are necessary to resolve lignocellulose-specific genes, which may exist at lower abundances or require specialized annotation.

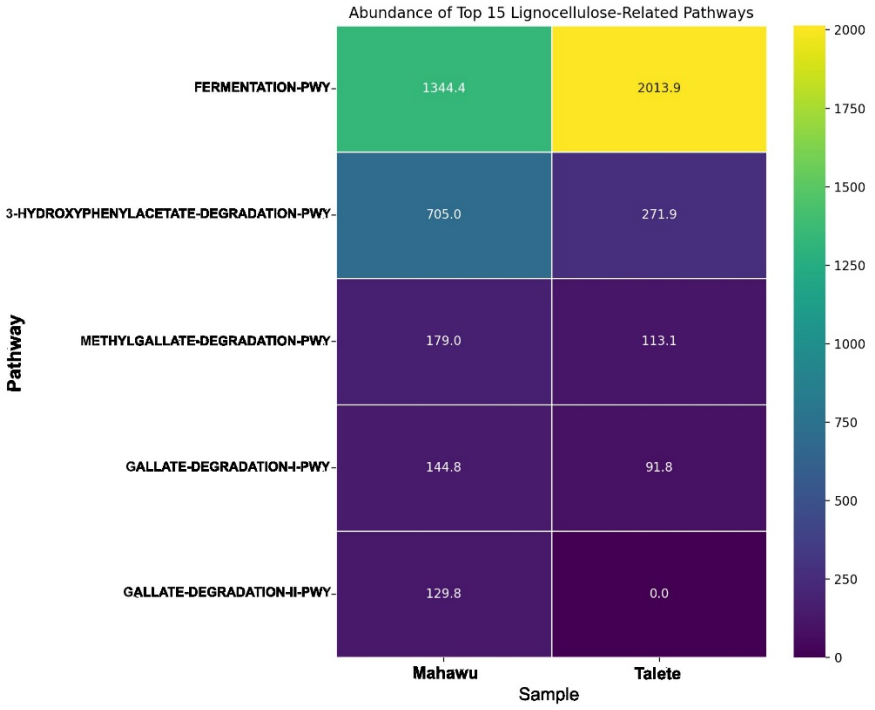

Fig. 10. KEGG-based PICRUSt analysis showing the relative abundance of the predicted functional lignocellulose-related gene content of microbial communities in Mahawu (MMF) and Talete (TCF). Pathway abundance values (yellow, green, blue, and purple) represent the number of genes from highest to lowest, respectively, and are normalized to the total number of genes present in a given pathway from each sample.

To resolve the similarities and differences in lignocellulose-specific pathways between the two forest types, a PICRUSt-based functional analysis was conducted. This involved filtering the original pathway abundance data using relevant keywords and assigning functional categories based on pathway names. Fig. 10 illustrates the abundance of the top 15 lignocellulose-related pathways in both locations.

The PICRUSt analysis focusing on lignocellulose-related pathways provides a detailed comparison of the functional potential of microbial communities in the Mahawu and Talete forests. The heatmap, illustrating the abundance of the top 15 lignocellulose-related pathways, reveals notable differences in the metabolic capabilities of these communities. Notably, the fermentation pathway (FERMENTATION-PWY) exhibits the highest abundance in both samples, with values of 1344.4 for Mahawu and 2013.9 for Talete. This substantial presence suggests that fermentation plays a key role in the degradation of complex plant materials, possibly contributing to the breakdown of lignocellulose into simpler compounds that can be further metabolized [40].

In contrast, the 3-hydroxyphenylacetate degradation pathway (3-HYDROXYPHENYLACETATE-DEGRADATION-PWY) shows a marked disparity between the two samples. Mahawu exhibits a relatively high abundance of 705.0, whereas Talete displays a significantly lower value of 271.9. This difference may indicate variations in the types of lignin-derived compounds in each forest environment or differences in the microbial populations responsible for their degradation. Similarly, the methylgallate degradation pathway (METHYLGALLATE-DEGRADATION-PWY) also demonstrates a notable distinction, with Mahawu showing an abundance of 179.0 compared to Talete's 113.1. These findings suggest that while both communities are capable of degrading methylgallate, Mahawu may harbor a more active or diverse set of microorganisms involved in this process.

Furthermore, the gallate degradation pathways (GALLATE-DEGRADATION-I-PWY and GALLATE-DEGRADATION-II-PWY) reveal notable patterns. In Mahawu, the abundances are 144.8 and 129.8 for the first and second pathways, respectively. However, Talete shows no activity for the second pathway (0.0), while the first pathway has a low abundance of 91.8. This pronounced difference may be attributed to the specific composition of lignocellulosic materials in each forest, which likely influences the activation of particular degradation pathways [41]. The absence of the second gallate degradation pathway in Talete may indicate either a lack of certain lignin-derived compounds or the presence of alternative degradation mechanisms not captured in this analysis.

Interpreting these findings in the context of metagenomic bioprospecting for lignocellulosic enzymes, it is evident that both Mahawu and Talete harbor microbial communities with distinct yet complementary capabilities for lignocellulose degradation. The high abundance of fermentation pathways in both samples emphasizes the importance of this process in breaking down complex plant materials [40]. Meanwhile, the variations observed in other degradation pathways highlight the potential for discovering novel enzymes or enzyme combinations adapted to specific lignocellulosic substrates. For instance, the higher abundance of 3-hydroxyphenylacetate and methylgallate degradation pathways in Mahawu may point to enzymes

effective in processing lignin-derived compounds, which are often recalcitrant to degradation.

3.5. Ecological and Functional Implications

The taxonomic and diversity differences between MMF and TCF reflect their respective ecological contexts. MMF's lower OTU richness (91 vs. 125) and more balanced taxonomic profile (e.g., 35% Proteobacteria, 10% Streptomycetaceae) suggest a stable microbial community adapted to consistent lignocellulosic inputs from native vegetation. Conversely, TCF's higher diversity (e.g., Shannon 4.44 vs. 4.11) and elevated abundances of Actinobacteria (30% vs. 25%) and Streptomycetaceae (15% vs. 10%) indicate a response to anthropogenic disturbances, such as organic matter variability, promoting a heterogeneous microbial assemblage. This aligns with studies reporting enhanced microbial diversity in disturbed soils [22].

Our findings are consistent with previous studies highlighting the dominance of Proteobacteria and Actinobacteriota in lignocellulose-rich environments [42,43]. However, the specific variations in phylum, class, order, family, and genus compositions between MMF and TCF emphasize the importance of considering site-specific ecological factors when analyzing microbial communities. For instance, the higher Actinobacteriota abundance in MMF aligns with studies emphasizing their role in the decomposition process in natural forests. Conversely, the enrichment of specific phyla in TCF, such as Bacteroidota and Gemmatimonadota, corresponds with findings from studies focused on high-moisture environments [44].

From a bioprospecting perspective, TCF's greater diversity and unique OTUs, along with higher abundances of lignocellulolytic genera such as *Streptomyces* (12% vs. 8%) and *Pseudomonas* (10% vs. 6%), suggest a richer reservoir of novel enzymes. These taxa are known for producing cellulases, xylanases, and ligninases [45], and their prominence in TCF may reflect adaptation to diverse substrates. Although MMF is less diverse, it hosts a robust core of degradative taxa (e.g., *Bacillus* at 3%), offering valuable insights into natural enzyme evolution. The absence of complete taxonomic exclusions across levels indicates a shared functional backbone, while the enhancements observed in TCF highlight its additional biotechnological potential.

Studies investigating tropical forest soils have also reported variations in microbial community structures driven by environmental conditions [46]. Our findings extend these observations by providing a comparative analysis of man-made versus natural forests, demonstrating the impact of anthropogenic activities on microbial biodiversity and functional potential. The identification of specific genera such as *Acidothermus*, *Bradyrhizobium*, and *Bacillus* as dominant players in these ecosystems highlights their potential for biotechnological applications, particularly in the production of lignocellulosic enzymes.

Interpreting the findings from the functional prediction using KEGG-based PICRUSt

analysis in the context of metagenomic bioprospecting for lignocellulosic enzymes reveals that both Mahawu and Talete harbor microbial communities with distinct yet complementary capabilities in lignocellulose degradation. The high abundance of fermentation pathways in breaking down complex plant materials. Meanwhile, variations in other degradation pathways emphasize the potential for discovering novel enzymes or enzyme combinations tailored to specific lignocellulosic substrates. For instance, the higher abundance of 3-hydroxyphenylacetate and methylgallate degradation pathways in Mahawu may facilitate the identification of enzymes effective in processing lignin-derived compounds, which are often recalcitrant to degradation.

Moreover, the complete absence of the second gallate degradation pathway in Talete presents an intriguing opportunity for comparative studies. Investigating the genetic and environmental factors underlying this difference could provide valuable insights into the adaptability and specialization of microbial communities across forest ecosystems. Such knowledge may inform strategies for optimizing lignocellulose degradation processes in industrial applications, including biofuel production and waste management.

4. Conclusions

This study provides the first comprehensive metagenomic characterization of bacterial communities in natural and man-made forest soils in North Sulawesi, Indonesia. The distinct microbial compositions between Mount Mahawu Forest and Tomohon City Forest highlight the impact of forest management on soil microbiomes. TCF supports a more diverse and taxonomically rich bacterial community, driven by anthropogenic influences, whereas MMF maintains a microbial profile reflective of natural forest conditions. These distinctions underscore both ecosystems as valuable reservoirs for lignocellulolytic enzymes, with TCF offering greater diversity and MMF providing evolutionary insights. The PICRUSt analysis of lignocellulose-related pathways offers a comprehensive view of the functional diversity within microbial communities from the Mahawu and Talete forests. The data reveal both shared and unique metabolic capabilities, highlighting the potential for discovering novel enzymes and understanding the ecological roles of these communities in lignocellulose degradation. These findings contribute to global soil microbiome databases and identify promising targets for bioprospecting, emphasizing the importance of preserving natural forest ecosystems. Future research should focus on functional metagenomics and broader geographical sampling to further explore the biotechnological potential of tropical soil microbiomes.

Abbreviations

NGS	next generation sequencing
rRNA	ribosomal ribonucleic acid
OTU	operational taxonomic unit

KEGG	kyoto encyclopedia of genes and genomes
PICRUSt	phylogenetic investigation of communities by reconstruction of unobserved states

Data availability statement

Data will be shared upon request by the readers.

Author's contributions

Feky Recky Mantiri: Writing – Original draft, Conceptualization, Methodology, Investigation, Resources, Formal analysis, Data curation, Writing – review & editing, Validation. **Carla Felly Kairupan**: Formal analysis, Validation, Data curation, Conceptualization, Supervision Writing – review & editing. **Sri Sudewi**: Conceptualization, Data curation, Supervision, Writing – review & editing. **Vic Axel Daniel Mantiri**: Formal analysis, Investigation, Data curation, Validation.

Declaration of Competing Interest

The authors declare no competing interest.

Acknowledgement

This study was funded by Sam Ratulangi University Grant Number: SP DIPA-23.17.2.677519/2023.

References

[1] Yadav A, Sharma V, Tsai ML, Chen CW, Sun PP, Nargotra P, et al. Development of lignocellulosic biorefineries for the sustainable production of biofuels: Towards circular bioeconomy. Bioresour Technol 2023;381:129145. https://doi.org/10.1016/J.BIORTECH.2023.129145.

[2] Berlemont R, Martiny AC. Genomic potential for polysaccharide deconstruction in bacteria. Appl Environ Microbiol 2015;81:1513–9. https://doi.org/10.1128/AEM.03718-14/SUPPL_FILE/ZAM999116043SO1.PDF.

[3] Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS. Microbial Cellulose Utilization: Fundamentals and Biotechnology. Microbiology and Molecular Biology Reviews 2002;66:506–77. https://doi.org/10.1128/MMBR.66.3.506-577.2002/ASSET/6F6A28CF-2C07-40BC-B7A7-BAF0F705A33E/ASSETS/GRAPHIC/MR032001411T.JPEG.

[4] Mukherjee G, Dhiman G, Akhtar N. Efficient Hydrolysis of Lignocellulosic Biomass: Potential Challenges and Future Perspectives for Biorefineries. Environmental Science and Engineering (Subseries: Environmental Science) 2017:213–37. https://doi.org/10.1007/978-3-319-48439-6_17.

[5] Baldrian P. Microbial activity and the dynamics of ecosystem processes in forest soils. Curr Opin Microbiol 2017;37:128–34. https://doi.org/10.1016/J.MIB.2017.06.008.

[6] Osorio-González CS, Chaali M, Hegde K, Brar SK, Kermanshahipour A, Avalos-Ramírez A. Production and Processing of the Enzymes from Lignocellulosic Biomass. Green Energy and Technology 2020:221–43. https://doi.org/10.1007/978-3-030-38032-8_11.

[7] Singhania RR, Ruiz HA, Awasthi MK, Dong C Di, Chen CW, Patel AK. Challenges in cellulase bioprocess for biofuel applications. Renewable and Sustainable Energy Reviews 2021;151:111622. https://doi.org/10.1016/J.RSER.2021.111622.

[8] Jan U, Feiwen R, Masood J, Chun SC. Characterization of Soil Microorganism from Humus and Indigenous Microorganism Amendments. Mycobiology 2020;48:392–8. https://doi.org/10.1080/12298093.2020.1816154.

[9] Verma N, Kumar V, Bansal MC. Valorization of Waste Biomass in Fermentative Production of Cellulases: A Review. Waste Biomass Valorization 2021;12:613–40. https://doi.org/10.1007/S12649-020-01048-8/METRICS.
[10] Soni SK, Sharma A, Soni R. Cellulases: Role in Lignocellulosic Biomass Utilization. Methods in Molecular Biology 2018;1796:3–23. https://doi.org/10.1007/978-1-4939-7877-9_1.
[11] Nargotra P, Sharma V, Lee YC, Tsai YH, Liu YC, Shieh CJ, et al. Microbial Lignocellulolytic Enzymes for the Effective Valorization of Lignocellulosic Biomass: A Review. Catalysts 2022;13:83. https://doi.org/10.3390/CATAL13010083.
[12] Yadav S, Reddy B, Dubey SK. De novo genome assembly and comparative annotation reveals metabolic versatility in cellulolytic bacteria from cropland and forest soils. Funct Integr Genomics 2020;20:89–101. https://doi.org/10.1007/S10142-019-00704-0/METRICS.
[13] Kukkar D, Sharma PK, Kim KH. Recent advances in metagenomic analysis of different ecological niches for enhanced biodegradation of recalcitrant lignocellulosic biomass. Environ Res 2022;215:114369. https://doi.org/10.1016/J.ENVRES.2022.114369.
[14] López-Mondéjar R, Zühlke D, Becher D, Riedel K, Baldrian P. Cellulose and hemicellulose decomposition by forest soil bacteria proceeds by the action of structurally variable enzymatic systems. Scientific Reports 2016;6:1–12. https://doi.org/10.1038/srep25279.
[15] Nadhifah H, Rahmani N, Mangunwardoyo W, Yopi, Atikana A, Ratnakomala S, et al. Xylanopectinolytic enzymes by marine actinomycetes from sediments of Sarena Kecil, North Sulawesi: high potential to produce galacturonic acid and xylooligosaccharides from raw biomass. Journal of Genetic Engineering and Biotechnology 2023;21:31. https://doi.org/10.1186/S43141-023-00488-8.
[16] Handayani I, Saad H, Ratnakomala S, Lisdiyanti P, Kusharyoto W, Krause J, et al. Mining indonesian microbial biodiversity for novel natural compounds by a combined genome mining and molecular networking approach. Mar Drugs 2021;19:316. https://doi.org/10.3390/MD19060316/S1.
[17] Straw CM, Henry GM, Love K, Carrow RN, Cline V. Evaluation of Several Sampling Procedures for Spatial Analysis of Natural Turfgrass Sports Field Properties. J Test Eval 2018;46:714–29. https://doi.org/10.1520/JTE20160467.
[18] Thijs S, De Beeck MO, Beckers B, Truyens S, Stevens V, Van Hamme JD, et al. Comparative evaluation of four bacteria-specific primer pairs for 16S rRNA gene surveys. Front Microbiol 2017;8:251189. https://doi.org/10.3389/FMICB.2017.00494/BIBTEX.
[19] Bellemain E, Carlsen T, Brochmann C, Coissac E, Taberlet P, Kauserud H. ITS as an environmental DNA barcode for fungi: An in silico approach reveals potential PCR biases. BMC Microbiol 2010;10:1–9. https://doi.org/10.1186/1471-2180-10-189/TABLES/4.
[20] Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 2011;17:10. https://doi.org/10.14806/EJ.17.1.200.
[21] Yu H, Chen Z, Wan Y, Sun X. Temperature-humidity-density dependent evaporation behaviour of clay and sandy clay. Eur J Soil Sci 2024;75:e13484. https://doi.org/10.1111/EJSS.13484.
[22] Hartmann M, Niklaus PA, Zimmermann S, Schmutz S, Kremer J, Abarenkov K, et al. Resistance and resilience of the forest soil microbiome to logging-associated compaction. ISME J 2014;8:226–44. https://doi.org/10.1038/ISMEJ.2013.141.
[23] Lladó S, López-Mondéjar R, Baldrian P. Forest Soil Bacteria: Diversity, Involvement in Ecosystem Processes, and Response to Global Change. Microbiology and Molecular Biology Reviews 2017;81. https://doi.org/10.1128/MMBR.00063-16/ASSET/F14C000C-59F6-4A9A-B45C-DB319A8584A8/ASSETS/GRAPHIC/ZMR0021724600003.JPEG.
[24] Zhu K, Liu T, Liu J, Cao X, Liu J, Wang J. Microbial degradation of lignocellulose. AIP Conf Proc 2021;2350. https://doi.org/10.1063/5.0048528/731276.

[25] Xue P, Minasny B, McBratney A, Pino V, Fajardo M, Luo Y. Distribution of soil bacteria involved in C cycling across extensive environmental and pedogenic gradients. Eur J Soil Sci 2023;74:e13337. https://doi.org/10.1111/EJSS.13337.

[26] Nikitina EP, Buyantueva LB, Abidueva EY, Sun CH. Taxonomic and ecophysiological characteristics of actinobacteria in soils of the dry steppe zone of the Selenga Highlands (Western Transbaikalia). Vavilovskii Zhurnal Genet Selektsii 2023;27:411–20. https://doi.org/10.18699/VJGB-23-49.

[27] Mersinkova Y, Yemendzhiev H, Kowalski A. Identification and Characterization of Natural Habitats of Electrochemically Active Bacteria. J Adv Biol Biotechnol 2020;23:19–25. https://doi.org/10.9734/JABB/2020/V23I130135.

[28] Nguyen VH, Wemheuer B, Song W, Bennett H, Palladino G, Burgsdorf I, et al. Functional characterization and taxonomic classification of novel gammaproteobacterial diversity in sponges. Syst Appl Microbiol 2023;46:126401. https://doi.org/10.1016/J.SYAPM.2023.126401.

[29] Saranraj P, Sayyed RZ, Sivasakthivelan P, Kokila M, Al-Tawaha ARM, Amala K, et al. Symbiotic Effectiveness of Rhizobium Strains in Agriculture. Plant Growth Promoting Microorganisms of Arid Region 2023:389–421. https://doi.org/10.1007/978-981-19-4124-5_18.

[30] Huang Z, Liu R, Chen F, Lai Q, Oren A, Shao Z. Nitrogeniibacter aestuarii sp. nov., a Novel Nitrogen-Fixing Bacterium Affiliated to the Family Zoogloeaceae and Phylogeny of the Family Zoogloeaceae Revisited. Front Microbiol 2021;12:755908. https://doi.org/10.3389/FMICB.2021.755908/BIBTEX.

[31] Norashirene MJ, Amin AA, Norhidayah D, Fithriah MAN. Identification of cellulolytic thermophiles based on 16S rDNA gene amplification analysis. CHUSER 2012 - 2012 IEEE Colloquium on Humanities, Science and Engineering Research 2012:413–8. https://doi.org/10.1109/CHUSER.2012.6504349.

[32] Kikuchi S, Fujitani H, Ishii K, Isshiki R, Sekiguchi Y, Tsuneda S. Characterisation of bacteria representing a novel Nitrosomonas clade: Physiology, genomics and distribution of missing ammonia oxidizer. Environ Microbiol Rep 2023;15:404–16. https://doi.org/10.1111/1758-2229.13158.

[33] Sly LI, Arunpairojana V, Hodgkinson MC. Pedomicrobium manganicum from Drinking-Water Distribution Systems with Manganese-Related “Dirty Water” Problems. Syst Appl Microbiol 1988;11:75–84. https://doi.org/10.1016/S0723-2020(88)80051-1.

[34] Zhu X, Fan F, Qiu H, Shao M, Li D, Yu Y, et al. New xylose transporters support the simultaneous consumption of glucose and xylose in Escherichia coli. MLife 2022;1:156–70. https://doi.org/10.1002/MLF2.12021.

[35] Cavaliere P, Brier S, Filipenko P, Sizun C, Raynal B, Bonnete F, et al. The stress sigma factor of RNA polymerase RpoS/σS is a solvent-exposed open molecule in solution. Biochemical Journal 2018;475:341–54. https://doi.org/10.1042/BCJ20170768.

[36] Zhong R, Cui D, Richardson EA, Phillips DR, Azadi P, Lu G, et al. Cytosolic Acetyl-CoA Generated by ATP-Citrate Lyase Is Essential for Acetylation of Cell Wall Polysaccharides. Plant Cell Physiol 2020;61:64–75. https://doi.org/10.1093/PCP/PCZ178.

[37] Li M, Xu X, Zou X, Hazelbauer GL. A Selective Tether Recruits Activated Response Regulator CheB to Its Chemoreceptor Substrate. MBio 2021;12. https://doi.org/10.1128/MBIO.03106-21/SUPPL_FILE/MBIO.03106-21-ST001.DOCX.

[38] Kabir M, Arif M, Ali F, Ahmad S, Swati ZNK, Yu DJ. Prediction of membrane protein types by exploring local discriminative information from evolutionary profiles. Anal Biochem 2019;564–565:123–32. https://doi.org/10.1016/J.AB.2018.10.027.

[39] Corkey BE, Deeney JT. The Redox Communication Network as a Regulator of Metabolism. Front Physiol 2020;11:567796. https://doi.org/10.3389/FPHYS.2020.567796/BIBTEX.

[40] Jardine KJ, McDowell N. Fermentation-mediated growth, signaling, and defense in plants. New Phytologist 2023;239:839–51. https://doi.org/10.1111/NPH.19015.

[41] Mikwa J, Gossens R, Defourny P. Forest degradation, a methodological approach using remote sensing techniques: A review. International Journal of Innovation and Scientific Research 2016;24:161–78. https://www.researchgate.net/publication/304299267_Forest_degradation_a_methodological_approach_using_remote_sensing_techniques_A_review.

[42] Houfani AA, Tláskal V, Baldrian P, Hahnke RL, Benallaoua S. Actinobacterial Strains as Genomic Candidates for Characterization of Genes Encoding Enzymes in Bioconversion of Lignocellulose. Waste Biomass Valorization 2022;13:1523–34. https://doi.org/10.1007/S12649-021-01595-8/METRICS.

[43] Georgiadou DN, Avramidis P, Ioannou E, Hatzinikolaou DG. Microbial bioprospecting for lignocellulose degradation at a unique Greek environment. Heliyon 2021;7. https://doi.org/10.1016/J.HELIYON.2021.E07122/ATTACHMENT/980CA4E1-822A-4A63-9DDB-12E91867EF33/MMC3.DOCX.

[44] Mujakić I, Piwosz K, Koblížek M. Phylum Gemmatimonadota and Its Role in the Environment. Microorganisms 2022;10:151. https://doi.org/10.3390/MICROORGANISMS10010151/S1.

[45] Nawaz MZ, Shang H, Sun J, Geng A, Ali SS, Zhu D. Genomic insights into the metabolic potential of a novel lignin-degrading and polyhydroxyalkanoates producing bacterium Pseudomonas sp. Hu109A. Chemosphere 2023;310:136754. https://doi.org/10.1016/J.CHEMOSPHERE.2022.136754.

[46] Ma S, Chen G, Tang W, Xing A, Chen X, Xiao W, et al. Inconsistent responses of soil microbial community structure and enzyme activity to nitrogen and phosphorus additions in two tropical forests. Plant Soil 2021;460:453–68. https://doi.org/10.1007/S11104-020-04805-9/METRICS.