diversity

MDPI

Article

De Novo Metagenomic Analysis of Microbial Community Contributing in Lignocellulose Degradation in Humus Samples Harvested from Cuc Phuong Tropical Forest in Vietnam

Thi-Thu-Hong Le [1,2], Thi-Binh Nguyen [1,2,3], Hong-Duong Nguyen [1], Hai-Dang Nguyen [1], Ngoc-Giang Le [1], Trong-Khoa Dao [1], Thi-Quy Nguyen [1], Thi-Huyen Do [1,2] and Nam-Hai Truong [1,2,*]

[1] Institute of Biotechnology, Vietnam Academy of Science and Technology, 18-Hoang Quoc Viet, Cau Giay, Hanoi 10072, Vietnam; lethuhong@ibt.ac.vn (T.-T.-H.L.); ntbinh@daihocthudo.edu.vn (T.-B.N.); duongnguyen96uet@gmail.com (H.-D.N.); nhdang1998@gmail.com (H.-D.N.); giangln@gmail.com (N.-G.L.); khoadt2103@gmail.com (T.-K.D.); quynhungcuong@yahoo.com (T.-Q.N.); dohuyen@ibt.ac.vn (T.-H.D.)

[2] Graduate University of Science and Technology, Vietnam Academy of Science and Technology, 18-Hoang Quoc Viet, Cau Giay, Hanoi 10072, Vietnam

[3] Faculty of Natural Sciences, Hanoi Metropolitan University (HMU), Hanoi 11300, Vietnam

* Correspondence: tnhai@ibt.ac.vn; Tel.: +84-24-3791-7980

Abstract: We aimed to investigate the microbial diversity, mine lignocellulose-degrading enzymes/proteins, and analyze the domain structures of the mined enzymes/proteins in humus samples collected from the Cuc Phuong National Park, Vietnam. Using a high-throughput Illumina sequencer, 52 Gbs of microbial DNA were assembled in 2,611,883 contigs, from which 4,104,872 open reading frames (ORFs) were identified. Among the total microbiome analyzed, bacteria occupied 99.69%; the five ubiquitous bacterial phyla included Proteobacteria, Bacteroidetes, Actinobacteria, Firmicutes, and Acidobacteria, which accounted for 92.59%. Proteobacteria (75.68%), the most dominant, was 5.77 folds higher than the second abundant phylum Bacteroidetes (13.11%). Considering the enzymes/proteins involved in lignocellulose degradation, 22,226 ORFs were obtained from the annotation analysis using a KEGG database. The estimated ratio of Proteobacteria/Bacteroidetes was approximately 1:1 for pretreatment and hemicellulases groups and 2.4:1 for cellulases. Furthermore, analysis of domain structures revealed their diversity in lignocellulose-degrading enzymes. CE and PL were two main families in pretreatment; GH1 and GH3-FN3 were the highest domains in the cellulase group, whereas GH2 and GH43 represented the hemicellulase group. These results validate that natural tropical forest soil could be considered as an important source to explore bacteria and novel enzymes/proteins for the degradation of lignocellulose.

Keywords: Cuc Phuong humus; Illumina de novo sequencing; lignocellulose degradation enzymes; DNA metagenome; tropical forest sample; white-rot fungi

Citation: Le, T.-T.-H.; Nguyen, T.-B.; Nguyen, H.-D.; Nguyen, H.-D.; Le, N.-G.; Dao, T.-K.; Nguyen, T.-Q.; Do, T.-H.; Truong, N.-H. De Novo Metagenomic Analysis of Microbial Community Contributing in Lignocellulose Degradation in Humus Samples Harvested from Cuc Phuong Tropical Forest in Vietnam. *Diversity* **2022**, *14*, 220. https://doi.org/10.3390/d14030220

Academic Editor: Ipek Kurtboke

Received: 10 February 2022
Accepted: 14 March 2022
Published: 17 March 2022

1. Introduction

Lignocellulose, which is composed of celluloses, hemicelluloses, and lignin, is derived from numerous sources that include agricultural crops and forest residues, along with bioenergy crops and forest products [1]. Lignocellulose, an abundant, sustainable, and renewable biomass, is used for the production of biofuels and other valuable products [2]. Biomass, an environmentally friendly entity, represents an inexpensive alternative to depleted fossil fuel resources. Hence, it can decrease global climate change and contribute to a sustainable and greener future [3]. Thus, the degradation of lignocellulose has gathered considerable attention from scientists and governments worldwide [4–6]. The hydrolysis of lignocellulose by chemical and physical methods effectively breaks down the rigid bonds. Nevertheless, this usually generates secondary products that are possibly environmentally hazardous, which inhibit consequent steps in the process [7,8]. Therefore, the approaches

based on using enzymes/proteins for lignocellulose degradation, which provide several advantages, such as low energy requirements and reduction in environmental pollution, are currently being selected for greener technology that produces second-generation biofuels [9]. However, effective breakdown of lignocellulose to desired products with complex reaction chains involving hydrolytic enzymes is rather challenging. Thus, exploring novel enzymes and microorganisms that exhibit desired characteristics for the degradation of plant biomass is currently an important issue. Among the microorganisms, bacteria are promising candidates that upgrade the feasibility of lignocellulose conversion for novel conversion strategies.

Soil is one of the most important sources for exploring new enzyme and microbiota candidates for effective lignocellulose degradation. Furthermore, it is a potential and complex ecosystem with diverse bacteria that play an important role in this environment. Environmental conditions such as geographical location and natural selection pressure also influence the biodiversity of living microbiota and their enzyme properties [10,11]. Hence, novel enzymes and bacteria suitably applied during harsh conditions such as pH, temperature, and salinity in the lignocellulose conversion process, are being continuously reported from soil analyses [12]. Certain studies have also demonstrated the diverse features of ubiquitous microorganisms present in the microbial community of each soil type, which suggests the adaptability of the microorganisms to a specific forest condition [12,13]. Cuc Phuong, located in Ninh Binh province, is the largest nature reserve in Vietnam. The park is one of the principal biodiversity sites in Vietnam; however, its microbial diversity remains unknown.

By comparing the culture method and metagenomics technology, studies have revealed that more than 99% of microorganisms have been unculturable [14]. Based on culture-independent high-throughput sequencing, we can potentially identify uncultivable microbiota and investigate the microbial community and taxonomic diversity at a high resolution. Consequently, a comprehensive determination of the microbial diversity involved in biomass degradation can be fundamental in evaluating the potential sources of novel enzymes and activities [15,16]. To date, various soil types have been analyzed to identify microbial communities promoting lignocellulose degradation [12,17–21]. In this study, we aimed to investigate the microbial diversity, mine lignocellulose-degrading enzymes/proteins, and analyze their domain structures in humus samples harvested from the surroundings of white-rot fungi prevailing deadwood site in Cuc Phuong tropical forest.

2. Materials and Methods

2.1. Sampling and Extraction of Metagenomics DNA

Humus was sampled at sites (GPS at 20.27776; 105.71137, within 10 km radius) in Cuc Phuong. Cuc Phuong National Park is located in Ninh Binh province, in the region of the Red River Delta in Vietnam. Furthermore, Cuc Phuong Park is the largest nature reserve and one of the primary biodiversity sites in Vietnam. The annual average temperature in Cuc Phuong is 20.6 °C. The annual humidity and precipitation are 90% and 2138 mm, respectively. Forty-five humus samples were collected from the degraded deadwood points with the growth of white-rot fungi (Figure 1). Each sample was taken about 100 g surrounding white-rot fungi. The pH values of all samples range between 6.9 and 7.3. After collection, the humus samples were transferred into an ice box and then stored at 4 °C.

The humus samples were pooled and homogenized in PBS buffer (137 mM NaCl, 2.7 mM KCl, 1.4 mM KH_2PO_4, and 10 mM Na_2HPO_4, pH 7.4). Consequently, the samples were centrifuged to remove the impurities, once at low-speed (500 rpm) for 10 min and then twice at 600 rpm for 10 min. Thereafter, microorganisms in the samples were harvested by centrifuging at 5000 rpm for 1 min. The obtained pellets were suspended in PBS buffer (pH 7.4) supplemented with 20% glycerol and stored at −80 °C.

Figure 1. Some pictures at the humus sample collection points in Cuc Phuong tropical forest.

Metagenomic DNA was extracted from bacterial samples prepared from 10 g humus. Each sample was suspended in 20 mL lysis buffer, containing 100 mM Tris, 100 mM EDTA, 100 mM Na_2HPO_4, 1.5 M NaCl, and 1% CTAB, (pH 8) with 0.1 mg/mL protease K and incubated at 37 °C for 30 min with gentle shaking. After incubation, the sample was treated with 3 mL 20% SDS and continually incubated at 65 °C for 30 min. Subsequently, the supernatant was collected by centrifuging the samples at 7000 rpm for 5 min. Then, phenol/chloroform/isoamyl alcohol (25:24:1, *v*/*v*) mixture was added to purify the DNA samples. The upper aqueous layer was harvested after centrifugation at 6500 rpm for 10 min at 4 °C. The DNA sample was precipitated using 6 mL isopropanol, followed by centrifugation at 13,000 rpm for 10 min. The DNA pellet was washed with 70% ethanol and resuspended in 300 µL of distilled water. The extracted DNA was verified using agarose gel electrophoresis, and the DNA concentration and quality were measured using a Nanophotometer P330 (Implen GmbH, Munich, Germany). Three preparations were combined together, and the mixed metagenomics DNA had a concentration of 113.25 µg/µL with $A_{260/280}$ and $A_{260/230}$ values of 1.925 and 2.235, respectively. In addition, the contamination by the inhibitors of DNA polymerase in the sample was assessed using PCR by amplifying 16S rDNA. A total DNA sample of approximately 100 µg was dispatched to BGI-Hong Kong Co. Ltd. for deep metagenome sequencing.

2.2. Metagenomic Sequencing

The metagenomic DNA from the humus samples was sequenced using Illumina HiSeq 2500 (Illumina, San Diego, CA, USA) to generate paired-end reads of 150 bps. The raw sequence data were filtered by SOAPnuke to remove noise, which was described as follows: reads containing 5% or more ambiguous base (N base); read containing adapter sequences (default: 15 bases overlapped by reads and adapter); reads containing 50% or lower quality (Q < 20) base. Subsequently, filtered reads were assembled de novo with both softwares IDBA (version 1.1.0) [22] http://i.cs.hku.hk/_alse/hkubrg/projects/idba_ud/, accessed on 3 August 2019 and MEGAHIT (version 1.0) softwares [23] https://github.com/voutcn/megahit, accessed on 3 August 2019 with a series of different k-mer sizes in parallel. The optimal k-mer was used to assemble the clean reads to contigs. Then, the clean reads were mapped to the assembled contigs through Bowtie 2 [24] with parameters "-p8–very-sensitive-local-k 100–score-min L,0,1.2" to get the coverage information for revising the contigs. MetaGeneMark (version 2.10, default parameters) was used to predict genes from assembled contigs [25]. The predicted genes were clustered using CD-HIT [26], with

a sequence identity threshold of 95% and an alignment coverage threshold of 90% [27]. The metagenomics sequences are available in the sequence read archive (SRA) under the accession number PRJNA715592.

2.3. Taxonomic Assignment

Taxonomic assignment of genes was performed with Blastp by aligning them against the NR database (accessed on 9 August 2019). Analysis of Nr BLAST output files was performed using the program MEGAN (version 4.6) [26]. This software reads the results of a BLAST comparison as input and attempts to place each read on a node in the NCBI taxonomy using the LCA algorithm [26]. The NCBI taxonomy was displayed as a tree, and the size of each node was scaled to indicate how many reads have been assigned to the corresponding taxon. The relative abundance of each taxonomy level from the same taxonomy was summed. The taxonomic level correlation was drawn using the Krona complement tool in Excel.

2.4. Functional Annotation

All the predicted genes were blasted against public databases, including SwissProt, KEGG (Kyoto Encyclopedia of Genes and Genomes) [28], EggNOG (Evolutionary genealogy of gene: Non-supervised Orthologous Groups, Version: 3.0) [29], and Nr (non-redundant protein sequence database) with e-value lower than 10^{-5} [26], and retrieved proteins with the highest sequence similarity with given genes along with their functional protein annotations.

2.5. Mining Genes Encoding Lignocellulose-Degrading Enzymes

In this study, we focused on the investigation of data obtained from the KEGG database. Compared to the other data, genes for lignocellulases were dominantly defined in KEGG. Therefore, genes encoding lignocellulose-degrading enzymes/proteins were initially mined according to KEGG's functional annotation and specially assigned according to Enzyme Commission number (EC) [28]. Based on annotated KEGG data, we obtained all the ORFs encoding enzymes/proteins related to the pretreatment and hydrolysis of lignocellulose. The code of the ORFs was linked with the total taxonomic profile of the humus to generate a combined function of lignocellulose degradation and taxonomy from the Nr and SwissProt database. In addition, the amino acid sequences of the lignocellulase were investigated for the presence of domain structures using Pfam (http://pfam.xfam.org, accessed on 22 December 2020) and HMMER from dbCAN databases [30] (Figure 2).

Moreover, the metagenome data were annotated using the Pfam database containing multiple alignments and hidden Markov model-based profiles (profiles HMMs) of complete protein domains [31]. Profile HMMs of available domains related to lignocellulose-degrading enzymes (including pretreatment enzymes, such as lytic polysaccharide monooxygenase) from the Pfam database were collected. Using the collection of profile HMMs, the homologous sequences were scanned against the metagenomic data sequences. The scan was executed by HMMER software (version 3.1) with the criteria as follows: the domain-based score value no less than 30, profile coverage greater than 0.5, and bias/score ratio less than 0.1 (Figure 2).

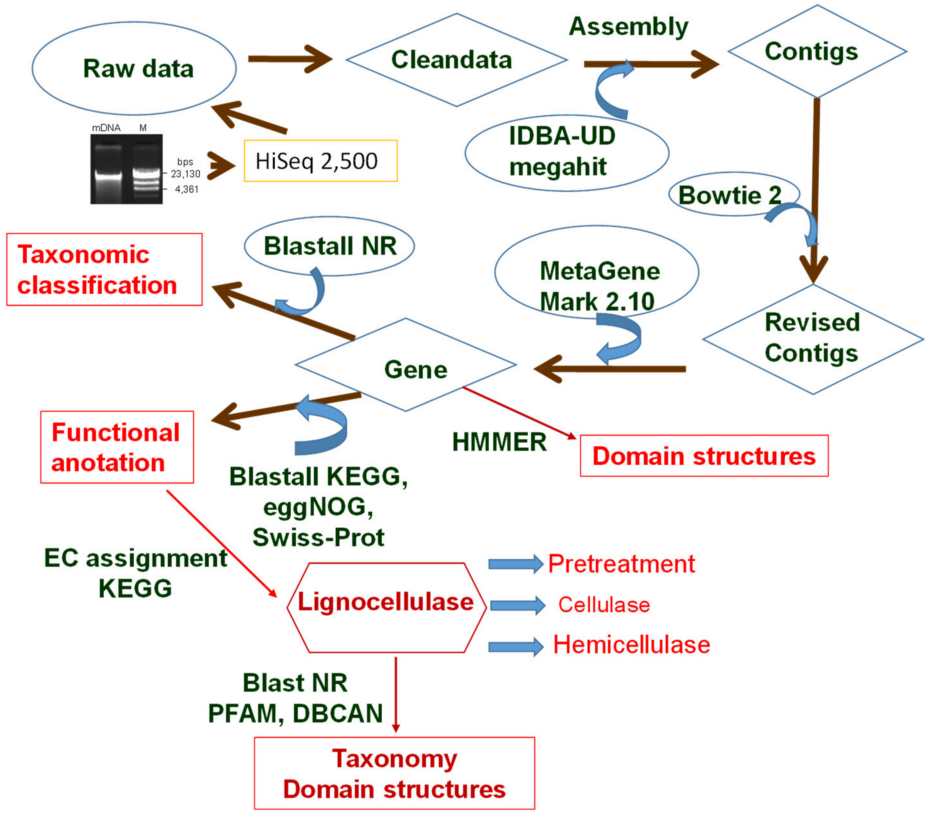

Figure 2. Workflow diagram for mining lignocellulose-degrading enzymes.

3. Results

3.1. Metagenome Sequencing of Cuc Phuong Tropical Forest Soil

The metagenomic DNA of bacteria in the humus collected from the Cuc Phuong tropical forest sequenced using the Illumina HiSeq platform was applied to assess the diversity of the microbial community and the potential proteins/enzymes involved in biomass degradation. From approximately 100 µg of metagenomic DNA, we obtained 345,471,086 high-quality clean reads and approximately 52 Gbs clean base. The assembly of clean data yielded 2,611,883 contigs with a total length of 2346 Mbs. The longest contig was 611,845 bps, and the N50 contig size and average contig were 1117 and 898 bps, respectively. Based on the assembly data, MetaGeneMark identified 4,104,872 protein-coding genes equivalent to 2074 Mbs. The N50, average, and the maximum gene lengths were 615, 505, and 20,541 bps, respectively (Table 1).

Table 1. Summary of metagenomic data obtained from the humus of Cuc Phuong tropical forest using Illumina HiSeq platform.

Category	Metric
Total reads	345,471,086
Total base (bp)	51,820,662,900
Number of contigs	2,611,883
Contig N50 (bp)	1117
Average contig length (bp)	898
Maximum contig length (bp)	611,845
Gene number	4,104,872
Gene N50 (bp)	615
Average gene length (bp)	505
Maximum gene length (bp)	20,541

3.2. Taxonomic Composition of Microbial Community in Cuc Phuong's Soil

From the 51.82 Gbs obtained from the metagenomic DNA data of Cuc Phuong tropical forest humus, which surrounded white-rot fungi that strongly degrade fallen forest trees, 4,104,872 genes were identified to encode proteins, among which 3,923,046 genes (equivalent to 95.57%) were annotated in the Nr database. The genes were classified using MEGAN (version 4.6) analysis, where 3,896,881 genes were assigned to bacteria, archaea, eukaryotes, and viruses. Bacteria were dominant with 3,884,879 ORFs accounting for 99.69% of the total identified ORFs, while the remaining belonged to archaea with 293 genes, eukaryotes with 1144 genes, and 10,565 genes representing the virus. The genes of the bacterial kingdom were affiliated to 111 phyla, 83 classes, 170 orders, 406 families, 1971 genera, with only 738 classified as species (Table 2).

Table 2. Overall analysis of humus bacterial metagenome by Nr BLAST in NCBI taxonomy database using the MEGAN (version 4.6).

Sources	Gene Num	Percentage (%)	Phylum	Class	Order	Family	Genus	Species
Bacteria	3,884,879	99.69	111	83	170	406	1971	738
Archaea	293	0.01	9	12	18	23	50	8
Eukaryota	1144	0.03	7	26	46	79	113	86
Viruses	10,565	0.27	0	0	2	14	101	84
Sum	3,896,881	100	131	118	237	523	2240	916

For deeper bacterial analysis, 93.29% of total genes were identified at the phylum level. Notably, Proteobacteria was the most abundant bacterial phylum, with 3,106,400 identified genes accounting for 75.68%, followed by Bacteroidetes at 13.11%. In addition, Actinobacteria, Firmicutes, and Acidobacteria accounted for 1.6%, 1.4%, and 0.8%, respectively. Thus, the five dominating phyla totally accounted for 92.59%, and the number of predicted genes originating from Proteobacteria was 5.77 folds higher than that of Bacteroidetes. The high abundance of these bacteria indicates that these phyla are important and play key roles in the humus bacterial community.

Similarly, analysis at class level showed that 93.68% of gene number was identified, and the most dominant class was Gammaproteobacteria (61.70%), followed by Betaproteobacteria (11.35%) and Alphaproteobacteria (6.85%) that belonged to Proteobacteria phylum. The next two classes included Sphingobacteria (6.39%) and Flavobacteriia (5.45%), which belonged to Bacteroidetes phylum. The other classes had a low ratio, below 1%. Three dominant orders comprised of Pseudomonadales (29.16%), Enterobacterales (22.26%), Burkholderiales (11.19%), followed by Sphingobacteriales (6.39%), Xanthomonadales (5.88%), Flavobacteriales (5.44%), Sphingomonadales (3.40%), Rhizobiales (2.66%), and Alteromonadales (1.68%), with others below 1%. The three highest families included Pseudomonadaceae (16.3%), Enterobacteriaceae (14.44%), and Moraxellaceae (11.02%). For genus level, only 45.27% of total genes were classified, and the ratio of all genera was lower than 10%. Species level was also investigated; however, the classification results were insignificant, with only 0.55% of defined genes, which did not complement the largely identified proportion (99.45%) (Figure 3, Table S1 in Supplementary Materials).

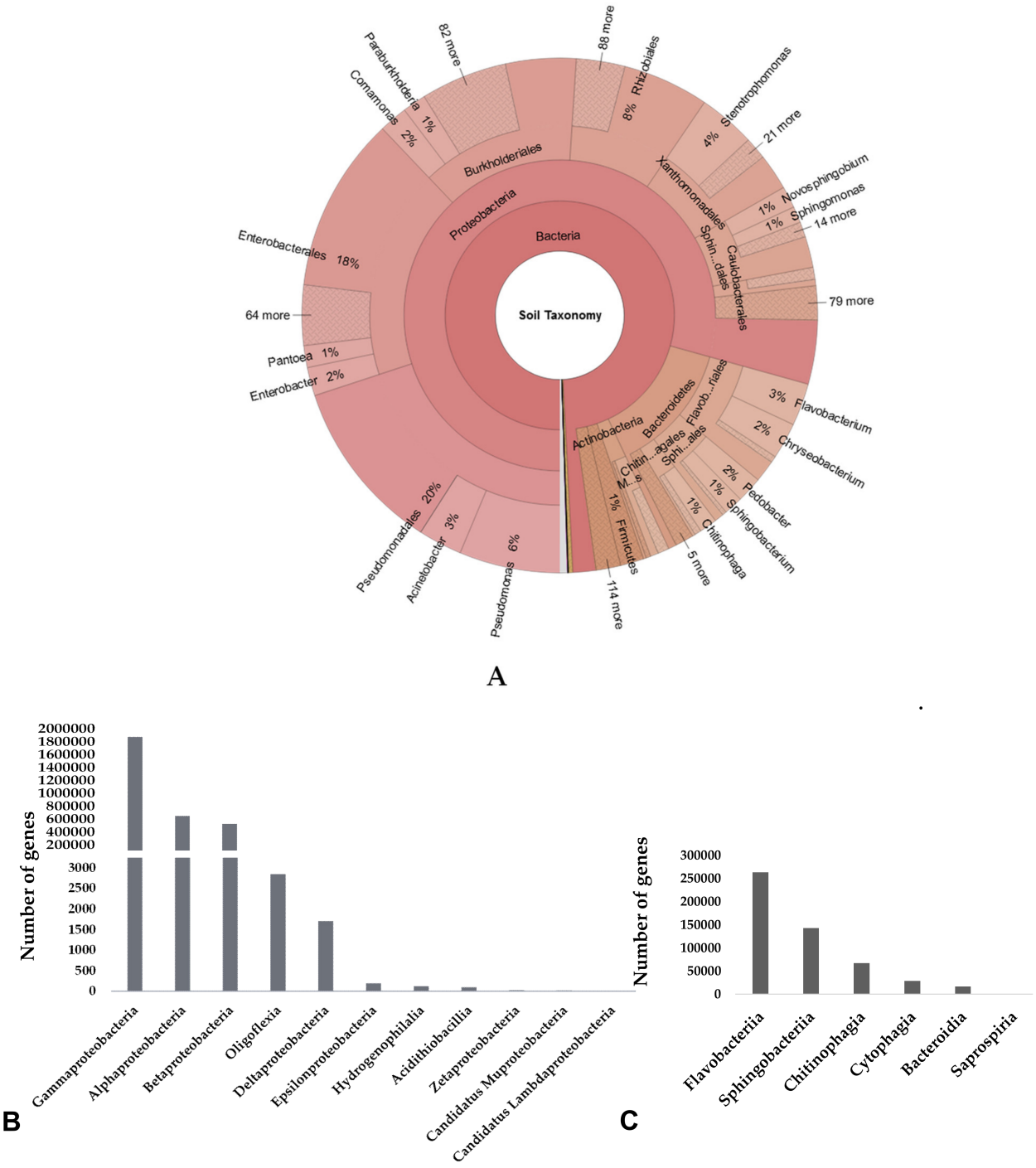

Figure 3. (**A**) Microbial community structure in humus samples surrounding white-rot fungi degrading deadwood in the Cuc Phuong forest at kingdom, phylum, order, and genus levels; (**B**) Proteobacteria composition at class levels; (**C**) Bacteroidetes composition at class levels.

3.3. Functional Profile of DNA Metagenome from Cuc Phuong's Humus

In order to obtain additional information to assess the functional potential associated with the microbial community, a set of metagenomic DNA from the humus samples was applied in databases such as Nr BLAST, SwissProt, KEGG, and eggNOG. Among the 4,104,872 identified ORFs, 3,925,740 ORFs corresponded to 95.64% of the total genes and were predicted to have functional annotation in at least one database (Table 3).

Table 3. Summary of functional annotation results.

	Total	Nr	Swissprot	KEGG	eggNOG	Overall
ORFs	4,104,872	3,923,046	2,382,630	2,809,791	3,279,853	3,925,740
%	100%	95.57%	58.04%	68.45%	79.90%	95.64%

Data obtained from the KEGG database were further investigated, and the pathway results were summarized into categories: cellular processes (cluster I), environmental information processing (cluster II), genetic information processing (cluster III), human diseases (cluster IV), metabolism (cluster V). Metabolism was the most dominant, which related to the growth of the microbial community, representing approximately 70% of the total defined ORFs in KEGG, where carbohydrate metabolism was observed to have 297,103 ORFs (Figure 4).

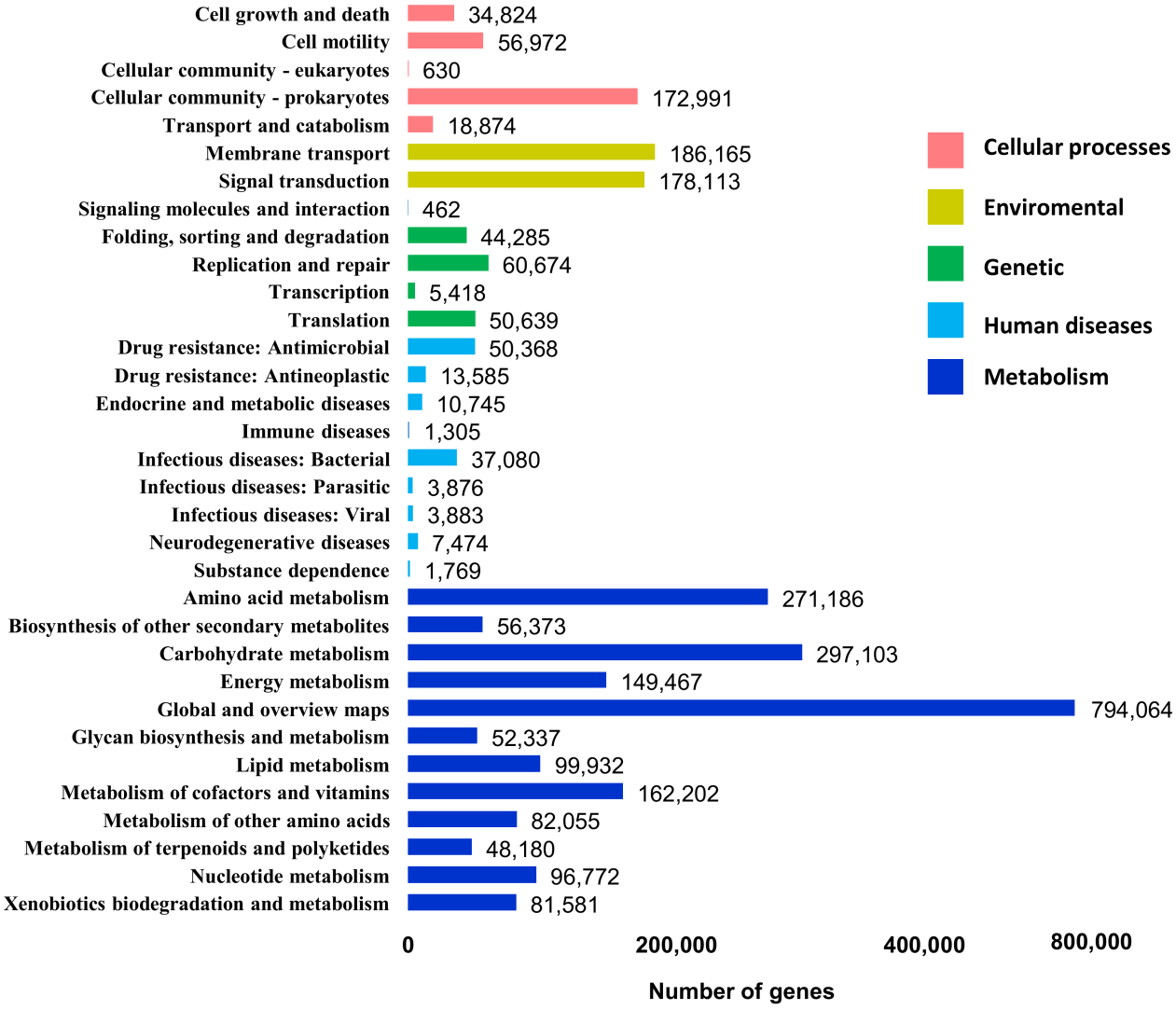

Figure 4. Summary of KEGG annotation. X-axis represents the number of genes that annotated each pathway, and y-axis lists annotated pathways in the particular subclass.

3.4. Putative Lignocellulose-Degrading Protein Encoding Genes

Among the ORFs related to carbohydrate metabolism based on the KEGG database, 22,226 ORFs encoding enzymes/proteins involved in lignocellulose degradation were annotated. Furthermore, 907 ORFs found for pretreatment were divided into four subgroups, which included pectin esterase, feruloyl esterase, laccase, and expansin, whereas ORFs

represented members of the other groups such as lignin peroxidase, lytic polysaccharide monooxygenase, and manganese peroxidase were absent. Pectin esterase was the most abundant group, with 815 sequences accounting for approximately 90% of genes. For enzymes belonging to cellulase, we mined 8301 sequences that involved five EC groups arranged in the order of high to low abundance as follows: β-glucosidase, endoglucanase, 6-phospho-beta-glucosidase, licheninase, cellobiohydrolase, and cellobiose phosphorylase, wherein β-glucosidase occupied more than 50%, which corresponded to 4272 defined sequences. Cellobiose dehydrogenase was absent in this data. For the hemicellulase group, 13,018 defined sequences were divided into 20 EC groups. Xyloglucan-active β-D-galactosidase was the most prominent group with 3288 ORFs, followed by α-L-fucosidase with 2279 ORFs, α-galactosidase with 1033 ORFs, α-L-arabinofuranosidase with 1016 ORFs, and others below 1000 ORFs per group. In addition, other ECs for hemicellulase, including acetyl xylan esterase, acetyl mannan esterase, α-D-xylosidase, α-L-fucosidase, were absent. All enzymes annotated in KEGG for lignocellulose degradation are listed in Table 4.

Table 4. Summary of ORFs encoding lignocellulases mined from the humus metagenomics DNA using KEGG database, Pfam, and HMMER (dbcan).

Cat *	Enzyme Name	ORF Number			Number of Complete ORFs Containing
	(EC . . .)	Total	Com **	Dom ***	Domain/Domain Types
P	Pretreatment	907	216	198	198/19 types
P1	Pectinesterase (EC 3.1.1.11)	815	199	181	61/CE8; 37/DUF4861; 29/CE2; 23/PL10; 15/Abhydrolase_3; 16/11 others
P4	Feruloylesterase (EC 3.1.1.73)	75	12	12	9/DUF3237; 3/Tannase
P3	Laccase (EC 1.10.3.2)	10	5	5	5/Cu3-Cu0-Cu2
P4	Expansin	7	0	0	
C	Cellulase	8301	1279	1058	1058/81 types
C1	β-glucosidase (EC 3.2.1.21)	4272	503	475	220/GH3-FN3; 93/GH1; 29/FN3; 29/GH3; 20/GH43; 11/GH3-Exop_C; 10/DUF4886; 10/CE3; 53/19 others
C2	Endoglucanase (EC 3.2.1.4)	2216	548	367	105/GH8; 72/GH5; 38/PeptidaseM42; 18 GH5-CBM6; 14/DUF285; 13/GH18; 10/CE2; 97/43 others
C3	6-phospho-beta-glucosidase (EC 3.2.1.86)	1718	213	210	152/GH1; 58/GH4
C4	Cellobiohydrolase (EC 3.2.1.91)	73	15	6	1/Alginate_lyase; 1/Amidase 3; 1/CBM2; 1/CBP_BcsO; 1/GH128 + Laminin G3; 1/Znribbon8
C5	Cellobiose phosphorylase (EC 2.4.1.20)	22	0	0	
H	Hemicellulase	13,018	2087	1828	1828/151 types
H1	xyloglucan-active β-D-galactosidase (EC 3.2.1.23)	3288	330	298	123/GH2; 28/GH42; 21/GH35; 20/DUF302; 18/GH43; 14/GH2 + CBM57; 13/Metallophos; 61/29 others
H2	α-L-fucosidase (EC 3.2.1.51)	2279	464	413	109/GH29; 81/GH95; 63/CE3; 46/Exo_endo_phos; 19/GH29 + CBM32; 16/CBM32; 13/GH33; 12/Big_2; 10/Abhydrolase_3; 9/GH117; 6/DUF1735 + CBM32; 29/19 others
H3	α-galactosidase (EC 3.2.1.22)	1033	163	134	65/GH36; 32/GH27; 16/GH4; 5/GH36 + GH27; 3/CBM51; 3/GH27 + CBM35; 2/Alginate_lyase; 8/8 others

Table 4. *Cont.*

Cat *	Enzyme Name	ORF Number			Number of Complete ORFs Containing
	(EC . . .)	Total	Com **	Dom ***	Domain/Domain Types
H4	α-L-arabinofuranosidase (EC 3.2.1.55)	1016	169	161	59/CE3; 47/GH51; 46/GH43; 4/GH43 + CBM32; 3/GH43 + GH121; 1/GH54; 1/Methyltransf-23
H5	endo-β-1,4 xylanase (EC 3.2.1.8)	885	230	175	65/Abhydrolase_3; 36/Peptidase_S9; 33/GH10; 15/CE15; 9/CBM6; 4/CE4; 13/9 others
H6	alpha-D-xylosidexylohydrolase (EC 3.2.1.177)	762	62	55	33/GH31; 9/GH31 + DUF5110; 6/Gal_mutarotas_2 + GH31; 2/DUF4968 + GH31 + DUF5110; 5/5 others
H7	xylan 1,4-beta-xylosidase (EC 3.2.1.37)	659	146	134	69/HTH_18; 45/GH43; 18/GH39; 2/AraC_binding + HTH_18
H8	Beta-mannosidase (EC 3.2.1.25)	611	46	37	22/GH2; 10/GH2 + Ig; 4/Ig; 1/GH158
H9	oligosaccharide reducing-end xylanase (EC 3.2.1.156)	552	100	73	31/GH43; 23/CHU; 4/GH8; 4/SprB; 3/CE4; 2/CBM9; 6/6 others
H10	β-mannanase (EC 3.2.1.78)	368	87	81	31/GH26; 17/GH5; 8/DUF1996; 7/GH44; 3/GH35; 3/CHU; 3/GH5 + CBM35; 9/9 others
H11	Endopolygalacturonaselyase (EC 4.2.2.2)	341	60	52	37/PL1; 3/PL1 + CBM77; 3/PL10; 3/PL2; 3/PL3; 2/CBM35 + PL1; 1/PL1 + LamininG3
H12	beta-fructofuranosidase (EC 3.2.1.26)	255	38	36	31/GH32; 2/Big_2; 1/CBM38 + GH32; 1/GH137; 1/PAN_4
H13	beta-D-glucuronidase (EC 3.2.1.31)	227	33	28	20/GH2; 6/GH141; 2/GH158
H14	Exopolygalacturonase (EC 3.2.1.67)	223	74	69	67/GH28; 2/NAD_binding_10
H15	Licheninase (EC 3.2.1.73)	175	52	52	48/GH16; 2/GH158 + GH16; 1/GH16 + CBM16; 1/GH16 + CBM6
H16	alpha-glucuronidase (EC 3.2.1.139)	161	17	16	16/GH67
H17	Exopolygalacturonaselyase (EC 4.2.2.9)	142	9	9	9/PL9
H18	Endopolygalacturonase (EC 3.2.1.15)	38	6	4	4/GH28
H19	endo-transglycosylase/hydrolase (EC 2.4.1.207)	2	1	1	1/GH16
H20	Acetylxylanesterase (EC 3.1.1.72)	1	0	0	

Cat *: catalog; Com **: complete ORFs; Dom ***: complete ORFs contain domain.

Among the 22,226 ORFs, which represented 0.54% of all ORFs involved in lignocellulose degradation, a substantial proportion of genes was assigned to a taxon by NCBI taxonomic classification, and only 107 (accounting for 0.49%) were not taxonomically classified. Among these, 22,092 classified ORFs (accounting for 99.39%) belonged to bacteria. Additionally, 28 phyla were identified in the lignocellulase data, which were dominated by Proteobacteria (11,288 ORFs, accounting for 50.79%), followed by Bacteroidetes (8164 ORFs, 36.73%), along with Firmicutes (3.43%), Actinobacteria (3.30%), Acidobacteria (1.99%), Verucomicrobia (0.53%), Cyanobacteria (0.11%), Planctomycetes (0.11%), and a total of

20 other phyla (0.22%) (Figure 5A, Table S2). The ratio of Bacteroidetes/Proteobacteria (0.72:1) was considerably higher than the ratio of the total bacterial structure in the humus (0.17:1). At the order level, Enterobacterales was identified as the most prominent order, accounting for 20.06%, followed by Flavobacteriales (15.14%) and Sphingobacteriales (11.62%).

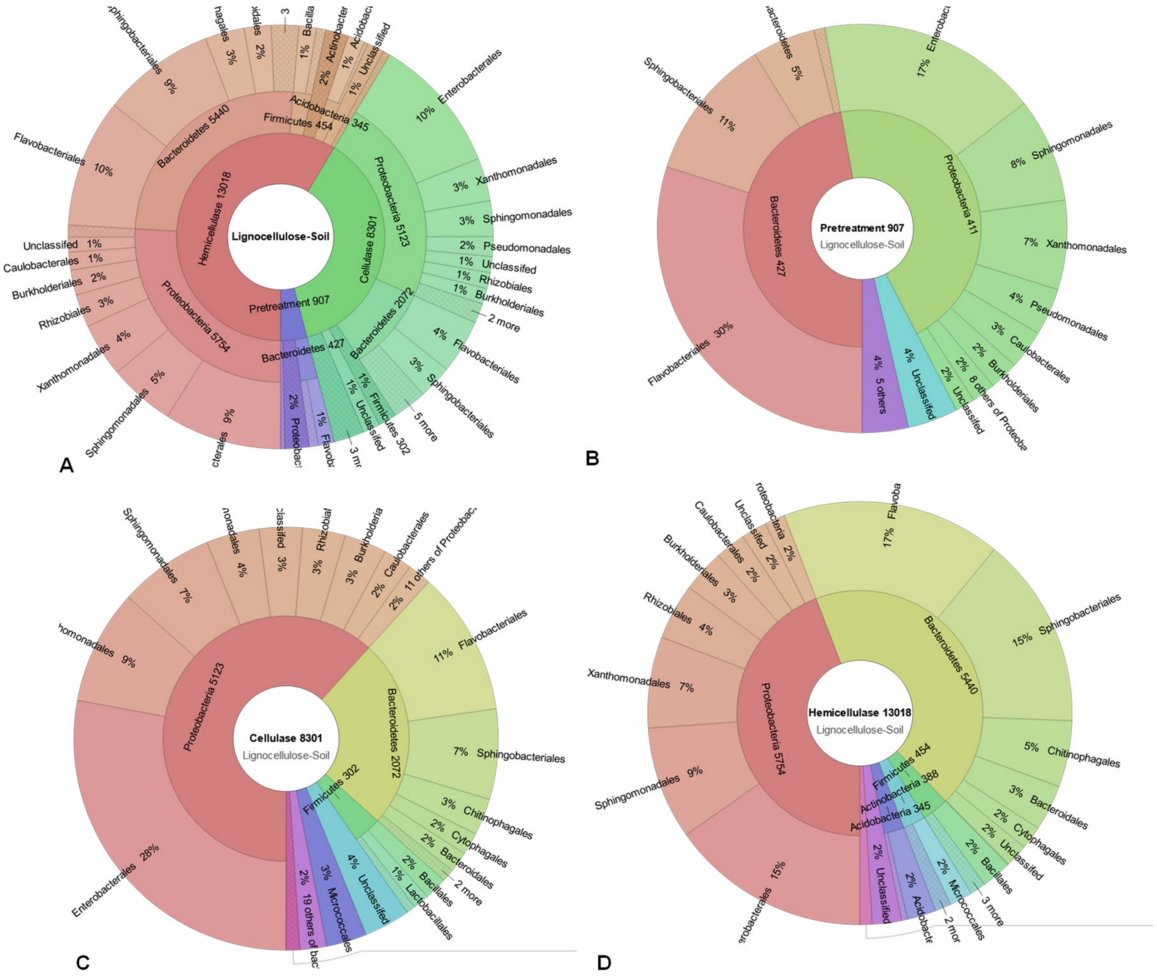

Figure 5. (**A**) Analysis of community structure of the humus bacteria harboring genes for lignocellulose degradation at phylum and order level, (**B**) lignocellulose pretreatment, (**C**) cellulase, and (**D**) hemicellulase. The number indicates the percentage and the number of genes.

For further analysis, in the pretreatment group, we observed that Bacteroidetes was the most abundant (427 ORFs, occupying 47.08%), slightly higher than Proteobacteia with 45.31%; whereas, for the hemicellulase group, Proteobacteria (44.20%) was slightly higher than Bacteroidetes (43.52%). For cellulase, the ratio of Proteobacteria and Bacteroidetes differed significantly, reaching 2.4 folds corresponding to Proteobacteria (61.72%) and Bacteroidetes (24.96%) individually. The Proteobacteria/Bacteroidetes ratio was 2.4, clearly indicating its predominance in pretreatment, and hemicellulose hydrolysis capacity revealed a Bacteroidetes/Proteobacteria ratio of approximately 1:1. Thus, Bacteroidetes appear to play a critical role in lignocellulose degradation.

A comparison of order taxonomy showed a notable difference from the total microbiota in the humus. Flavobacteriales, Sphingobacteriales, Enterobacterales with 29.88%, 19.63%, 17.42%, respectively, were three dominant orders in pretreatment; Enterobacterales (27.90%) and Flavobacteriales (11.02%) were two abundant orders in cellulase; whereas for hemicellulase, Flavobacteriales, Enterobacterales, and Sphingobacteriales were the

three dominant orders accounting for 16.72%, 15.26%, 14.65%, respectively. In the total humus microbiota, Sphingobacteriales (6.39%,) Xanthomonadales (5.88%), and Flavobacteriales (5.44%) represented the second abundant order cataloged below 10%; whereas, the most dominant orders were Pseudomonadales with 29.16%, followed by Enterobacterales (22.26%) and Burkholderiales (11.19%). In contrast, Pseudomonadales only accounted for 3.75%, 4.04%, and 0.45% representing the pretreatment, cellulase, and hemicelulase, respectively (Figure 5, Table S2). Thus, Pseudomonadales was the typical order in the humus sample but not for the lignocellulase in the humus. Enterobacterales dominated both the humus sample and lignocellulose-degrading enzyme/protein from the humus. Moreover, the order Flavobacteriales predominated all types of lignocellulase. Thus, Flavobacteriales and Enterobacterales orders belonging to Bacteroideles and Proteobacteria phyla, respectively, play an important role in lignocellulose degradation in the humus.

In order to gain further insight into the mechanism of lignocellulose degradation by the communities, we specifically observed the distribution of domains of the predicted lignocellulases using the Pfam and HMMER databases. Among the 22,226 ORFs encoding lignocellulases annotated in KEGG, only 3582 ORFs (16.12%) were complete, in which 3084 ORFs (equivalent to 86.1% of the completed ORFs) were assigned domains. For the pretreatment enzyme/protein group, 198 domain-containing complete ORFs were divided into 19 domain types. Carbohydrate Esterases (CE) and Polysaccharide Lyases (PL) were the main families that yielded approximately 62%; other domains such as abhydrolase, tannase, copper oxidase were also discovered. Family CE8 was the most predominant domain, accounting for 32% in pretreatment (Figure 6A, Table S3).

Similarly, the completed 1058 ORFs containing domains belonging to the cellulase group were defined into 81 domain types. Major domains were GH families that accounted for more than 80%, representing GH1 and GH3-FN3 with 245 ORFs and 220 ORFs, respectively, followed by GH8, GH5, and GH4. Furthermore, other groups such as peptidase M42, FN3, GH3, GH43 were also identified in the data. Comparing the enzymes annotated by KEGG and domains, GH3 was the predominant domain in enzyme β-glucosidase. In particular, GH3 collocated with the FN3 module, which assists in enzyme conformation and activity, was the most predominant domain with 220 ORFs in the enzyme group. In addition, GH8 was attributed to the endoglucanase group, followed by GH5 at the second level. In particular, the 6-phospho-beta-glucosidase group had only two types of domains present, which included GH1 and GH4. Moreover, combining domain and taxonomy showed that the GH1, GH8, and GH4 domains prevailed in Proteobacteria with 77%, 91%, and 95%, respectively. In contrast, GH3-FN3 and GH5 were equally divided in both Proteobacteria and Bacteroideles. Overall, in the taxonomy for domain-containing complete enzymes involving cellulose degradation, Proteobacteria dominated Bacteroideles and other phyla (Figure 6B, Table S4).

Additionally, our analysis showed that 151 domain types of 1828 completed ORFs were adopted from hemicellulase data, indicating the diversity of hemicellulase domains. In practically all domains, GH families were allocated diversely, which outperformed GH2 featured for xyloglucan-active β-D-galactosidase, followed by GH43 present in certain enzymes such as α-L-arabinofuranosidase, xylan 1,4-beta-xylosidase, oligosaccharide reducing-end xylanase, with the equivalent ratio in Bacteroidetes and Proteobacteria phyla; whereas CE3, abhydrolase, and GH28 were predominant in Proteobacteria phyla, and other abundant domains, such as GH29, GH95, were mainly present in Bacteroidetes (Figure 6C, Table S5).

Alternatively, using profile HMMs of lytic polysaccharide monooxygenase (LPMO) and multiple-copper oxidase (MCO) collected from Pfam to search and annotate the predicted ORFs in the metagenome data, we found 69 hits and 901 hits that belonged to LPMOs and MCOs, respectively. All LPMO hits contained the LPMO10 domain, whereas the MCO hits consisted of at least one or more of the four copper oxidase. Similarly, 224 ORFs were annotated as catalase/peroxidase, which enclosed the catalase domain, and 53 ORFs were defined as feruloyl esterase. Approximately half of the annotated hits were complete ORFs.

Notably, 37 MCO hits, which were annotated by profile HMMs, could not be annotated with NCBI, KEGG, SwissProt, eggNOG databases (Table 5).

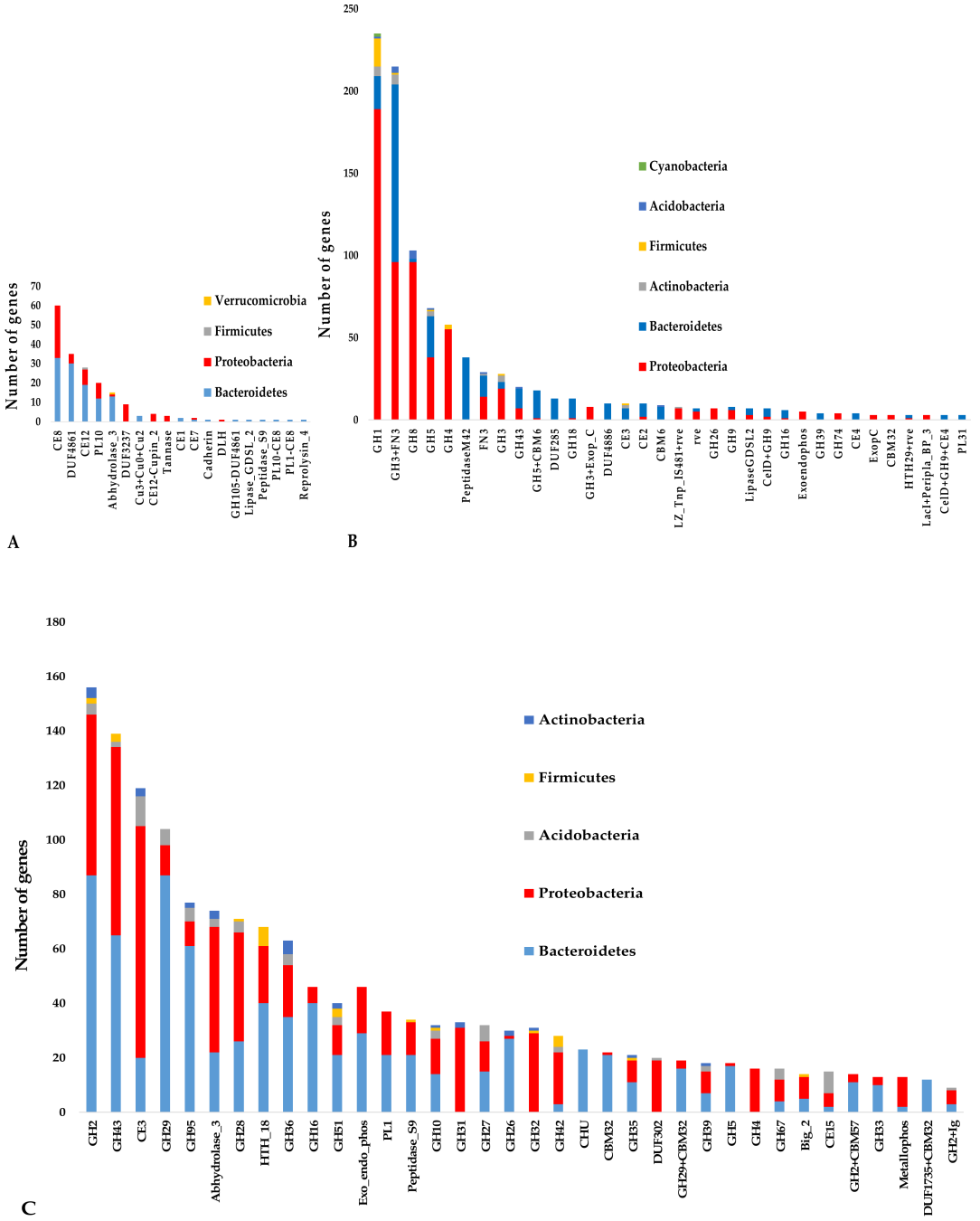

Figure 6. (**A**) Bacterial phyla possess domain-containing complete ORFs encoding protein/enzyme involved in lignocellulose pretreatment, (**B**) cellulose, and (**C**) hemicellulase. CE: carbohydrate esterase; PL: polysaccharide lyases; GDLS: motid Gly-Asp-Ser-Leu sequence; CBM: carbohydrate binding model; GH: glycoside hydrolase family; DUF: domain of unknown function; HTH: helix-turn-helix.

Table 5. Summary of ORFs encoding pretreatment enzymes mined from the humus metagenomics DNA using profile HMMs.

No	Enzyme Name (EC . . .)	Total	Complete	% Complete	Domain
1	Catalase/Peroxidase (EC 1.11.1.21)	224	142	63%	Catalase
2	Feruloylesterase (EC 3.1.1.73)	53	35	66%	Tannase
3	Multi-copper oxidase	901	483	54%	Cu-oxidase, Cu_oxidase_2, Cu_oxidase_3, Cu_oxidase_4
4	Lytic polysaccharide monooxygenase (EC 1.14.99.54)	69	33	48%	LPMO_10

4. Discussion

In most soils, common phyla that are usually abundant are Proteobacteria, Bacteroidetes, Acidobacteria, Actinobacteria, and Firmicutes [13,32]. The ubiquitously dominant phyla were significantly found in the biomass study of Arundo donax, Eucalyptus amaldulensis, and Populus nigra using high-throughput sequencing of the 16S rRNA gene [11]. Using metagenomics analysis, two soil samples near phosphate rock chemical plants in Shuangsheng revealed that the most dominant bacteria were Proteobacteria at 38.56% and 57.85% [33]. Other taxa, including Acidobacteria, Verrucomicrobia, Cyanobacteria, and Planctomycetes were also present in the samples. Furthermore, other findings revealed that the phyla Acidobacteria and Proteobacteria have a higher relative abundance in soil environments, such as forest soil [18,21], crop soil [19], and lettuce soil [20]. In litter and deadwood samples, soils usually contain abundant copiotrophic bacteria from phyla Proteobacteria and Bacteroidetes. The bacterial communities exhibit successional stages along the decay process in litter and wood [13]. Consistent with our study, the five phyla, Proteobacteria, Bacteroidetes, Actinobacteria, Firmicutes, and Acidobacteria, were predominant, which accounted for 92.59% of total phyla. It is noted that in our study Proteobacteria was the most dominant phylum accounting for 75.68%, especially the most abundant of Gammaproteobacteria class (61.70%) belonging to this phylum in the metagenome from humus samples of Cuc Phuong tropical forest. It has been known that in soil samples, pH is one of the major factors affecting the composition and diversity of the bacterial community [34–36]. Each group of bacteria usually grows at a narrow optimal pH range. Some studies showed that the abundance of Proteobacteria subgroups positively relates to neutral pH, and some Acidobacteria subphylums grow at acid pH range, whereas the effect of pH on Bacteroides abundance is not clear [34,37]. Perhaps Bacteroides growth relates to other factors such as nutrient composition than pH. In our humus samples, the range of pH was 6.9–7.3. Therefore, it is possible to be suitable for the most abundant of Proteobacteria subgroups, typically Gammaproteobacteria.

Certain studies also showed that Bacteroidetes usually represent about 10% of the total microbiota in soils [38]. Additionally, bacteria belonging to the phylum Bacteroidetes are known to contain several genes encoding the enzymes for polysaccharide degradation [39,40]. A study of high-Arctic peat soils of Svalbard showed that most genes assigned to lignocellulose degraders belonged to phyla Bacteroidetes, Actinobacteria, and Verrucomicrobia, representing approximately 70% of these genes [41]. In our results, Proteobacteria was present in the soil microbiota, but Bacteroidetes appear to play a vital role in lignocellulose degradation. The majority of identified enzymes/proteins involved in lignocellulose degradation were assigned to Proteobacteria and Bacteroidetes. The estimated ratio of Proteobacteria/Bacteroidetes harboring genes coding lignocellulose enzymes was 1:1 in pretreatment and hemicellulase groups and 2.4:1 in cellulose. In particular, the result showed that Flavobacteriales of phylum Bacteroidetes were one of the most dominant

orders harboring putative lignocellulose enzymes. Thus, both phyla contain members recognized for their role in the degradation, which were discovered in Cuc Phuong National Park humus that surrounded the sites prominent with white-rot fungi degrading deadwood were similar in microbial construction as observed in other soil investigations [11,42,43]. Other findings have described the lignocellulose degradation capabilities of bacterial strains belonging to Proteobacteria, Bacteroidetes, Actinobacteria, and Firmicutes [44–48]. Recently, the role of bacteria of phylum Bacteroidetes in polymeric carbon degradation in soils has been investigated using the KEGG database [17]. However, the analyzed results are noted to be highly dependent on the analytical method and the selected databases.

For enzyme classification and domain analysis, we identified several important enzymes involved in the degradation of three major components of lignocellulose, which include cellulose, hemicellulose, and lignin. For lignin degradation, CE and PL (815 of the 2,808,791 ORFs annotated by KEGG, accounting for 0.03%) were predominant domains, which were consistent with our previous data [49]. Additionally, we mined 10 genes encoding laccases, including 5 complete genes that encode proteins containing 3 domains of Cu3-Cu-Cu2 in the metagenomics data from the humus samples. Laccase is an important enzyme of the multicoperoxidase group involved in lignin degradation. Laccases are mainly found in fungi and plants, but few reports described bacterial laccases. Furthermore, studies showed that laccases are predominant in soil environments [12,18,50,51]. In previous investigations, this enzyme was not identified in our metagenomics data obtained from microbial termite's gut, goat's rumen [49,52]. The presence of laccase in the humus is probably due to deep sequencing and sufficient oxygen level available for the soil microbiota. However, the number of ORFs coding for enzymes/proteins involved in lignocellulose pretreatment in this study was lower than that of the other studies. For example, a study investigated bacterial genes coding for lignocellulases in soil samples collected from a tropical forest in the Luquillo Experimental Forest, Puerto Rico, by sequencing the metagenomic DNA using Roche 454 GS FLX Titanium technology (Branford, CT). Among 68,911 functional annotated genes, 3133 CE (accounted 4.48%), 413 lignin oxidases (0.59%), 282 PL (0.41%) and 240 lignin-degrading auxiliary oxidases (0.35%) were identified [53]. The low abundance of the lignolytic enzymes in our data may be attributed to the characteristics of the sample collected from the sites of white-rot fungi degrading deadwood. The composition of the bacterial community in deadwood can be considerably influenced by fungal communities [51]. In nature, white-rot fungi are the most efficient, completely degrading lignin into CO_2 and H_2O [54]. This ability of fungi is due to the secretion of extracellular enzymes, [55] including lignin peroxidase, manganese peroxidase, oxy oxidoreductase/laccase, and other accessory enzymes [56] when growing under limited nutrients (C:N ratio), especially in wood and soil [57].

Once degraded, cellulases and hemicellulases can be attached to the loosened lignocellulose. The overview of lignocellulases in this study is similar to the abundance patterns across the rain forest soil [53] and compost [58]. In soil, fungi normally degrade polymers into organic compounds of low molecular mass, which is preferentially used by bacteria [59,60]. In our data, GH1, GH8, and GH4 representing β-glucosidase, endoglucanase, and 6-phospho-β-glucosidase, respectively, were the most frequent typical enzyme families derived from the phylum Proteobacteria. In contrast, in other metagenomic data, GH5 affiliated with endoglucanase was dominant and belonged to phylum Bacteroidetes [49]. Notably, the enzymes were not collocated with other activity domains or non-catalytic accessory domains. These domains were perhaps featured for cellulose-degrading enzymes from soil's metagenome with the predominance of Proteobacteria and contained only one catalytic domain. Remarkably, the domain analysis showed that approximately 90% of the GH3 domain in β-glucosidase was associated with an accessory module FN3, and the construct was the most abundantly associated domain type, and the GH3-FN3 architecture was observed in both Protoeobacteria and Bacteroideles. Furthermore, the FN3 module, a non-catalytic domain, is known to decompress cellulose surface [61] and assists in enzyme conformation and activity [62–64]. This result was also consistent with previous

findings that β-glucosidase contains FN3 but not CBM [65,66]. Our previous investigation mined GH3 from goat rumen, in which the GH3 domain associated with the FN3 domain accounted for 90.9% [63].

This is the first study investigating the bacterial community and the diversity of lignocellulases from bacteria in humus samples surrounding white-rot fungi degrading deadwood in tropical forest Cuc Phuong, Vietnam. Interestingly, the bacterial community structure and abundance pattern of cellulases and hemicellulases in the humus were similar to the samples from the soil in the rainforest. Due to the enzymes secreted by white-rot fungi for lignin conversion, lignolytic enzymes in the humus were low in abundance.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/d14030220/s1, Table S1: Microbial community structure in humus samples surrounding white-rot fungi degrading deadwood in the Cuc Phuong forest at kingdom, phylum, order, and genus levels; Table S2: Analysis of community structure of the humus bacteria harboring genes for lignocellulose degradation at phylum and order level; Table S3: Bacterial phyla possess domain-containing complete ORFs encoding protein/enzyme involved in lignocellulose pretreatment; Table S4: Bacterial phyla possess domain-containing complete ORFs encoding cellulases; Table S5: Bacterial phyla possess domain-containing complete ORFs encoding hemicellulose.

Author Contributions: T.-T.-H.L. analyzed metagenomics deep sequencing data; prepared and revised the manuscript; T.-B.N. analyzed metagenomics deep sequencing data; H.-D.N. (Hong-Duong Nguyen) analyzed domains of proteins/enzymes by HMMER; H.-D.N. (Hai-Dang Nguyen) analyzed taxonomy; T.-K.D., N.-G.L. and T.-Q.N. collected samples and extracted metagenomic DNA of bacteria in humus sample; T.-K.D. analyzed domains of proteins/enzymes by HMMER; T.-H.D. analyzed metagenomics deep sequencing data; N.-H.T. is principal investigator of the project NĐT.50.GER/18; built a strategy for all activities of the project. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the grant from the Bilaterial International Project Metagen-Lig, Code: NĐT.50.GER/18, from Ministry of Science and Technology (MOST), Vietnam and Federal Ministry of Education and Research, Germany. The funders had no role in the design of the study, in the collection, analyses, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in the sequence read archive (SRA) under the accession number PRJNA715592.

Acknowledgments: Authors would like to express our sincere thanks to Juergen Pleiss (Institute of Biochemistry and Technical Biochemistry, University of Stuttgart, Germany) for providing conditions for our Ph.D. student T.-K. Dao to analyze metagenimc DNA sequences using HMMER software. We also acknowledge S.V.N. Vijayendra (Principal Scientist, Central Food Technological Research Institute-Resource Centre, Hyderabad−500007, India) for initial editing this manuscript. This work was performed at the National Key Laboratory of Gene Technology, Institute of Biotechnology, Vietnam Academy of Science and Technology (VAST), Vietnam.

Conflicts of Interest: The authors declare that they have no competing interests.

References

1. Sharma, H.K.; Xu, C.; Qin, W. Biological Pretreatment of Lignocellulosic Biomass for Biofuels and Bioproducts: An Overview. *Waste Biomass Valoriz.* **2019**, *10*, 235–251. [CrossRef]
2. Bhatia, L.; Sharma, A.; Bachheti, R.K.; Chandel, A.K. Lignocellulose derived functional oligosaccharides: Production, properties, and health benefits. *Prep. Biochem. Biotechnol.* **2019**, *49*, 744–758. [CrossRef] [PubMed]
3. Carriquiry, M.A.; Du, X.; Timilsina, G.R. Second generation biofuels: Economics and policies. *Energy Policy* **2011**, *39*, 4222–4234. [CrossRef]
4. Himmel, M.E.; Ding, S.Y.; Johnson, D.K.; Adney, W.S.; Nimlos, M.R.; Brady, J.W.; Foust, T.D. Biomass recalcitrance: Engineering plants and enzymes for biofuels production. *Science* **2007**, *315*, 804–807. [CrossRef] [PubMed]

5. Ragauskas, A.J.; Williams, C.K.; Davison, B.H.; Britovsek, G.; Cairney, J.; Eckert, C.A.; Frederick, W.J.; Hallett, J.P.; Leak, D.J.; Liotta, C.L.; et al. The path forward for biofuels and biomaterials. *Science* **2006**, *311*, 484–489. [CrossRef]
6. Ravindran, R.; Hassan, S.S.; Williams, G.A.; Jaiswal, A.K. A review on bioconversion of agro-industrial wastes to industrially important enzymes. *Bioengineering* **2018**, *5*, 93. [CrossRef]
7. Achinivu, E.C. Protic ionic liquids for lignin extraction—A lignin characterization study. *Int. J. Mol. Sci.* **2018**, *19*, 428. [CrossRef]
8. Kumari, D.; Singh, R. Pretreatment of lignocellulosic wastes for biofuel production: A critical review. *Renew. Sustain. Energy Rev.* **2018**, *90*, 877–891. [CrossRef]
9. Sharma, A.; Kumar, R.; Rathi, M.; Bhatia, D.; Malik, D.K. Lignocellulose biodegradation: An advance technology for sustainable environment. *Biosci. Biotechnol. Res. Commun.* **2018**, *11*, 634–637. [CrossRef]
10. Van den Burg, B. Extremophiles as a source for novel enzymes. *Curr. Opin. Microbiol.* **2003**, *6*, 213–218. [CrossRef]
11. Ventorino, V.; Aliberti, A.; Faraco, V.; Robertiello, A.; Giacobbe, S.; Ercolini, D.; Amore, A.; Fagnano, M.; Pepe, O. Exploring the microbiota dynamics related to vegetable biomasses degradation and study of lignocellulose-degrading bacteria for industrial biotechnological application. *Sci. Rep.* **2015**, *5*, 8161. [CrossRef] [PubMed]
12. López-Mondéjar, R.; Algora, C.; Baldrian, P. Lignocellulolytic systems of soil bacteria: A vast and diverse toolbox for biotechnological conversion processes. *Biotechnol. Adv.* **2019**, *37*, 107374. [CrossRef] [PubMed]
13. Lladó, S.; López-Mondéjar, R.; Baldrian, P. Forest Soil Bacteria: Diversity, Involvement in Ecosystem Processes, and Response to Global Change. *Microbiol. Mol. Biol. Rev.* **2017**, *81*, e00063-16. [CrossRef] [PubMed]
14. Rappé, M.S.; Giovannoni, S.J. The Uncultured Microbial Majority. *Annu. Rev. Microbiol.* **2003**, *57*, 369–394. [CrossRef]
15. Duan, C.J.; Feng, J.X. Mining metagenomes for novel cellulase genes. *Biotechnol. Lett.* **2010**, *32*, 1765–1775. [CrossRef]
16. Thompson, C.E.; Beys-da-Silva, W.O.; Santi, L.; Berger, M.; Vainstein, M.H.; Guimarães, J.A.; Vasconcelos, A.T.R. A potential source for cellulolytic enzyme discovery and environmental aspects revealed through metagenomics of Brazilian mangroves. *AMB Express* **2013**, *3*, 65. [CrossRef]
17. Alteio, L.V.; Schulz, F.; Seshadri, R.; Varghese, N.; Rodriguez-Reillo, W.; Ryan, E.; Goudeau, D.; Eichorst, S.A.; Malmstrom, R.R.; Bowers, R.M.; et al. Complementary Metagenomic Approaches Improve Reconstruction of Microbial Diversity in a Forest Soil. *mSystems* **2020**, *5*, e00768-19. [CrossRef]
18. Baldrian, P.; Kolaiřík, M.; Štursová, M.; Kopecký, J.; Valášková, V.; Větrovský, T.; Žifčáková, L.; Šnajdr, J.; Rídl, J.; Vlček, Č.; et al. Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. *ISME J.* **2012**, *6*, 248–258. [CrossRef]
19. Lopes, A.R.; Manaia, C.M.; Nunes, O.C. Bacterial community variations in an alfalfa-rice rotation system revealed by 16S rRNA gene 454-pyrosequencing. *FEMS Microbiol. Ecol.* **2014**, *87*, 650–663. [CrossRef]
20. Schreiter, S.; Ding, G.C.; Heuer, H.; Neumann, G.; Sandmann, M.; Grosch, R.; Kropf, S.; Smalla, K. Effect of the soil type on the microbiome in the rhizosphere of field-grown lettuce. *Front. Microbiol.* **2014**, *5*, 144. [CrossRef]
21. Turlapati, S.A.; Minocha, R.; Bhiravarasa, P.S.; Tisa, L.S.; Thomas, W.K.; Minocha, S.C. Chronic N-amended soils exhibit an altered bacterial community structure in Harvard Forest, MA, USA. *FEMS Microbiol. Ecol.* **2013**, *83*, 478–493. [CrossRef] [PubMed]
22. Peng, Y.; Leung, H.C.M.; Yiu, S.M.; Chin, F.Y.L. IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **2012**, *28*, 1420–1428. [CrossRef] [PubMed]
23. Li, D.; Liu, C.M.; Luo, R.; Sadakane, K.; Lam, T.W. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **2015**, *31*, 1674–1676. [CrossRef]
24. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [CrossRef] [PubMed]
25. Zhu, W.; Lomsadze, A.; Borodovsky, M. Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res.* **2010**, *38*, e132. [CrossRef]
26. Huson, D.H.; Auch, A.F.; Qi, J.; Schuster, S.C. MEGAN analysis of metagenomic data. *Genome Res.* **2007**, *17*, 377–386. [CrossRef]
27. Li, W.; Godzik, A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **2006**, *22*, 1658–1659. [CrossRef]
28. Kanehisa, M.; Araki, M.; Goto, S.; Hattori, M.; Hirakawa, M.; Itoh, M.; Katayama, T.; Kawashima, S.; Okuda, S.; Tokimatsu, T.; et al. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **2008**, *36*, D480–D484. [CrossRef]
29. Powell, S.; Szklarczyk, D.; Trachana, K.; Roth, A.; Kuhn, M.; Muller, J.; Arnold, R.; Rattei, T.; Letunic, I.; Doerks, T.; et al. eggNOG v3.0: Orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res.* **2012**, *40*, D284–D289. [CrossRef]
30. Mistry, J.; Finn, R.D.; Eddy, S.R.; Bateman, A.; Punta, M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* **2013**, *41*, e121. [CrossRef]
31. Mistry, J.; Chuguransky, S.; Williams, L.; Qureshi, M.; Salazar, G.A.; Sonnhammer, E.L.L.; Tosatto, S.C.E.; Paladin, L.; Raj, S.; Richardson, L.J.; et al. Pfam: The protein families database in 2021. *Nucleic Acids Res.* **2021**, *49*, D412–D419. [CrossRef]
32. Li, J.G.; Shen, M.C.; Hou, J.F.; Li, L.; Wu, J.X.; Dong, Y.H. Effect of different levels of nitrogen on rhizosphere bacterial community structure in intensive monoculture of greenhouse lettuce. *Sci. Rep.* **2016**, *6*, 25305. [CrossRef] [PubMed]
33. Feng, G.; Xie, T.; Wang, X.; Bai, J.; Tang, L.; Zhao, H.; Wei, W.; Wang, M.; Zhao, Y. Metagenomic analysis of microbial community and function involved in cd-contaminated soil. *BMC Microbiol.* **2018**, *18*, 11. [CrossRef] [PubMed]
34. Rousk, J.; Bååth, E.; Brookes, P.C.; Lauber, C.L.; Lozupone, C.; Caporaso, J.G.; Knight, R.; Fierer, N. Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J.* **2010**, *4*, 1340–1351. [CrossRef]

35. Lauber, C.L.; Hamady, M.; Knight, R.; Fierer, N. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl. Environ. Microbiol.* **2009**, *75*, 5111–5120. [CrossRef]
36. Cho, S.J.; Kim, M.H.; Lee, Y.O. Effect of pH on soil bacterial diversity. *J. Ecol. Environ.* **2016**, *40*, 10. [CrossRef]
37. Dimitriu, P.A.; Grayston, S.J. Relationship between soil properties and patterns of bacterial β-diversity across reclaimed and natural boreal forest soils. *Microb. Ecol.* **2010**, *59*, 563–573. [CrossRef] [PubMed]
38. Fierer, N. Embracing the unknown: Disentangling the complexities of the soil microbiome. *Nat. Rev. Microbiol.* **2017**, *15*, 579–590. [CrossRef]
39. Berlemont, R.; Martiny, A.C. Genomic potential for polysaccharide deconstruction in bacteria. *Appl. Environ. Microbiol.* **2015**, *81*, 1513–1519. [CrossRef]
40. Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P.M.; Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **2014**, *42*, D490–D495. [CrossRef]
41. Tveit, A.; Schwacke, R.; Svenning, M.M.; Urich, T. Organic carbon transformations in high-Arctic peat soils: Key functions and microorganisms. *ISME J.* **2013**, *7*, 299–311. [CrossRef]
42. Alessi, A.M.; Bird, S.M.; Bennett, J.P.; Oates, N.C.; Li, Y.; Dowle, A.A.; Polikarpov, I.; Young, J.P.W.; McQueen-Mason, S.J.; Bruce, N.C. Revealing the insoluble metasecretome of lignocellulose-degrading microbial communities. *Sci. Rep.* **2017**, *7*, 2356. [CrossRef] [PubMed]
43. Jiménez, D.J.; Chaves-Moreno, D.; Van Elsas, J.D. Unveiling the metabolic potential of two soil-derived microbial consortia selected on wheat straw. *Sci. Rep.* **2015**, *5*, 13845. [CrossRef] [PubMed]
44. Berlemont, R.; Allison, S.D.; Weihe, C.; Lu, Y.; Brodie, E.L.; Martiny, J.B.H.; Martiny, A.C. Cellulolytic potential under environmental changes in microbial communities from grassland litter. *Front. Microbiol.* **2014**, *5*, 639. [CrossRef] [PubMed]
45. Himmel, M.E.; Xu, Q.; Luo, Y.; Ding, S.Y.; Lamed, R.; Bayer, E.A. Microbial enzyme systems for biomass conversion: Emerging paradigms. *Biofuels* **2010**, *1*, 323–341. [CrossRef]
46. Koeck, D.E.; Pechtl, A.; Zverlov, V.V.; Schwarz, W.H. Genomics of cellulolytic bacteria. *Curr. Opin. Biotechnol.* **2014**, *29*, 171–183. [CrossRef] [PubMed]
47. López-Mondéjar, R.; Zühlke, D.; Větrovský, T.; Becher, D.; Riedel, K.; Baldrian, P. Decoding the complete arsenal for cellulose and hemicellulose deconstruction in the highly efficient cellulose decomposer Paenibacillus O199. *Biotechnol. Biofuels* **2016**, *9*, 104. [CrossRef]
48. Sukharnikov, L.O.; Cantwell, B.J.; Podar, M.; Zhulin, I.B. Cellulases: Ambiguous nonhomologous enzymes in a genomic perspective. *Trends Biotechnol.* **2011**, *29*, 473–479. [CrossRef]
49. Do, T.H.; Dao, T.K.; Nguyen, K.H.V.; Le, N.G.; Nguyen, T.M.P.; Le, T.L.; Phung, T.N.; van Straalen, N.M.; Roelofs, D.; Truong, N.H. Metagenomic analysis of bacterial community structure and diversity of lignocellulolytic bacteria in Vietnamese native goat rumen. *Asian-Australas. J. Anim. Sci.* **2018**, *31*, 738–747. [CrossRef]
50. Baldrian, P. Fungal laccases-occurrence and properties. *FEMS Microbiol. Rev.* **2006**, *30*, 215–242. [CrossRef]
51. Odriozola, I.; Abrego, N.; Tláskal, V.; Zrůstová, P.; Morais, D.; Větrovský, T.; Ovaskainen, O.; Baldrian, P. Fungal Communities Are Important Determinants of Bacterial Community Composition in Deadwood. *mSystems* **2021**, *6*, e01017-20. [CrossRef]
52. Do, T.H.; Nguyen, T.T.; Nguyen, T.N.; Le, Q.G.; Nguyen, C.; Kimura, K.; Truong, N.H. Mining biomass-degrading genes through Illumina-based de novo sequencing and metagenomic analysis of free-living bacteria in the gut of the lower termite Coptotermes gestroi harvested in Vietnam. *J. Biosci. Bioeng.* **2014**, *118*, 665–671. [CrossRef] [PubMed]
53. DeAngelis, K.M.; Gladden, J.M.; Allgaier, M.; D'haeseleer, P.; Fortney, J.L.; Reddy, A.; Hugenholtz, P.; Singer, S.W.; Vander Gheynst, J.S.; Silver, W.L.; et al. Strategies for enhancing the effectiveness of metagenomic-based enzyme discovery in lignocellulolytic microbial communities. *Bioenergy Res.* **2010**, *3*, 146–158. [CrossRef]
54. Rodríguez-Couto, S. Industrial and environmental applications of white-rot fungi. *Mycosphere* **2017**, *8*, 456–466. [CrossRef]
55. Wesenberg, D.; Kyriakides, I.; Agathos, S.N. White-rot fungi and their enzymes for the treatment of industrial dye effluents. *Biotechnol. Adv.* **2003**, *22*, 161–187. [CrossRef]
56. Ruiz-Dueñas, F.J.; Martínez, Á.T. Microbial degradation of lignin: How a bulky recalcitrant polymer is efficiently recycled in nature and how we can take advantage of this. *Microb. Biotechnol.* **2009**, *2*, 164–177. [CrossRef]
57. Kirk, T.K.; Farrell, R.L. Enzymatic "combustion": The microbial degradation of lignin. *Annu. Rev. Microbiol.* **1987**, *41*, 465–505. [CrossRef] [PubMed]
58. Allgaier, M.; Reddy, A.; Park, J.I.; Ivanova, N.; D'Haeseleer, P.; Lowry, S.; Sapra, R.; Hazen, T.C.; Simmons, B.A.; Vandergheynst, J.S.; et al. Targeted discovery of glycoside hydrolases from a switchgrass-adapted compost community. *PLoS ONE* **2010**, *5*, e8812. [CrossRef]
59. Žifčáková, L.; Větrovský, T.; Lombard, V.; Henrissat, B.; Howe, A.; Baldrian, P. Feed in summer, rest in winter: Microbial carbon utilization in forest topsoil. *Microbiome* **2017**, *5*, 122. [CrossRef]
60. Štursová, M.; Žifčáková, L.; Leigh, M.B.; Burgess, R.; Baldrian, P. Cellulose utilization in forest litter and soil: Identification of bacterial and fungal decomposers. *FEMS Microbiol. Ecol.* **2012**, *80*, 735–746. [CrossRef]
61. Kataeva, I.A.; Seidel, R.D.; Shah, A.; West, L.T.; Li, X.L.; Ljungdahl, L.G. The fibronectin type 3-like repeat from the Clostridium thermocellum cellobiohydrolase CbHa promotes hydrolysis of cellulose by modifying its surface. *Appl. Environ. Microbiol.* **2002**, *68*, 4292–4300. [CrossRef] [PubMed]

62. Ding, S.Y.; Xu, Q.; Crowley, M.; Zeng, Y.; Nimlos, M.; Lamed, R.; Bayer, E.A.; Himmel, M.E. A biophysical perspective on the cellulosome: New opportunities for biomass conversion. *Curr. Opin. Biotechnol.* **2008**, *19*, 218–227. [CrossRef] [PubMed]
63. Nguyen, K.H.V.; Dao, T.K.; Nguyen, H.D.; Nguyen, K.H.; Nguyen, T.Q.; Nguyen, T.T.; Nguyen, T.M.P.; Truong, N.H.; Do, T.H. Some characters of bacterial cellulases in goats' rumen elucidated by metagenomic DNA analysis and the role of fibronectin 3 module for endoglucanase function. *Asian-Australas. J. Anim. Sci.* **2021**, *34*, 867–879. [CrossRef] [PubMed]
64. Wilson, D.B. Three microbial strategies for plant cell wall degradation. *Ann. N. Y. Acad. Sci.* **2008**, *1125*, 289–297. [CrossRef]
65. Do, T.H.; Le, N.G.; Dao, T.K.; Nguyen, T.M.P.; Le, T.L.; Luu, H.L.; Nguyen, K.H.V.; Nguyen, V.L.; Le, L.A.; Phung, T.N.; et al. Metagenomic insights into lignocellulose-degrading genes through Illuminabased de novo sequencing of the microbiome in vietnamese native goats' rumen. *J. Gen. Appl. Microbiol.* **2018**, *64*, 108–116. [CrossRef]
66. Sweeney, M.D.; Xu, F. Biomass converting enzymes as industrial biocatalysts for fuels and chemicals: Recent developments. *Catalysts* **2012**, *2*, 244–263. [CrossRef]