J. Microbiol. Biotechnol. (2017), 27(9), 1670–1680
https://doi.org/10.4014/jmb.1705.05008

jmb

Metagenomic SMRT Sequencing-Based Exploration of Novel Lignocellulose-Degrading Capability in Wood Detritus from *Torreya nucifera* in Bija Forest on Jeju Island

Han Na Oh[1], Tae Kwon Lee[2], Jae Wan Park[1], Jee Hyun No[2], Dockyu Kim[3], and Woo Jun Sul[1]*

[1]*Department of Systems Biotechnology, Chung-Ang University, Anseong 17546, Republic of Korea*
[2]*Department of Environmental Engineering, Yonsei University, Wonju 26493, Republic of Korea*
[3]*Division of Life Sciences, Korea Polar Research Institute, Incheon 21990, Republic of Korea*

Received: May 4, 2017
Revised: June 12, 2017
Accepted: June 19, 2017

First published online
June 21, 2017

**Corresponding author*
Phone: +82-31-670-4707;
Fax: +82-31-670-3108;
E-mail: sulwj@cau.ac.kr

pISSN 1017-7825, eISSN 1738-8872

Lignocellulose, composed mostly of cellulose, hemicellulose, and lignin generated through secondary growth of woody plant, is considered as promising resources for biofuel. In order to use lignocellulose as a biofuel, biodegradation besides high-cost chemical treatments were applied, but knowledge on the decomposition of lignocellulose occurring in a natural environment is insufficient. We analyzed the 16S rRNA gene and metagenome to understand how the lignocellulose is decomposed naturally in decayed *Torreya nucifera* (L) of Bija forest (Bijarim) in Gotjawal, an ecologically distinct environment. A total of 464,360 reads were obtained from 16S rRNA gene sequencing, representing diverse phyla; Proteobacteria (51%), Bacteroidetes (11%) and Actinobacteria (10%). The metagenome analysis using single molecules real-time sequencing revealed that the assembled contigs determined originated from Proteobacteria (58%) and Actinobacteria (10.3%). Carbohydrate Active enZYmes (CAZy)- and Protein families (Pfam)-based analysis showed that Proteobacteria was involved in degrading whole lignocellulose, and Actinobacteria played a role only in a part of hemicellulose degradation. Combining these results, it suggested that Proteobacteria and Actinobacteria had selective biodegradation potential for different lignocellulose substrates. Thus, it is considered that understanding of the systemic microbial degradation pathways may be a useful strategy for recycle of lignocellulosic biomass, and the microbial enzymes in Bija forest can be useful natural resources in industrial processes.

Keywords: Lignocellulose degradation, Bija forest, metagenome, 16S rRNA, CAZy, Pfam

Introduction

Lignocellulose, a major compartment of plant dry matter, is the most promising feedstock as a renewable natural resource for biofuels [1]. Lignocellulose consists mostly of three components (*i.e.*, cellulose, hemicellulose, and lignin) contributing to the hydrolytic stability of the plant cell wall, which is resistant to deconstruction and hydrolysis [2]. The cross-linking of sugar monomers in hemicellulose and cellulose and of phenolic units in lignin leads to recalcitrance of this biomass [3]. Thus, it is important to understand the enzymatic hydrolysis process in order to transform (in terms of chemical composition or structure) lignocellulosic biomass into value-added products by chemical or biological approaches. High temperature and acidic conditions are expected to convert lignocellulosic biomasses into such degradation products as 5-hydroxymethyl furfural, xylose, sorbitol, succinic acid, and ethanol [4]. Complex processes like dilute-acid pretreatment, hydrothermal pretreatment, and liquid hot water pretreatment are applied to reduce lignocellulose recalcitrance and to solubilize each component in lignocellulose [5]. Lignocellulose transformation into value-added products can be a feasible means for reduction of greenhouse gas emissions and for increasing

the security of the energy supply. Recently, biological delignification has gained increasing research attention and importance because it involves more economical, eco-friendly, and less hazardous procedures than do the physicochemical treatments [6].

Thus, an understanding of microbial enzymatic mechanisms has provided insights into improvement of the efficiency of lignocellulosic biomass decomposition. To study microbial metabolic pathways of lignocellulosic biomass degradation and hydrolysis, metagenomic approaches have recently been applied in various environments, including sugarcane bagasse [7], peat swamp forest [8], wet tropical forest soils [9], and wood-feeding insects [10, 11]. These studies on short reads from the Illumina platform have revealed that various bacterial species and functional genes are strongly linked to degradation of lignocellulose at different temperatures, pH, and oxygen availability levels. Nonetheless, short-read sequencing could not directly encompass entire regions of lignocellulosic biomass degradation operons in the microbial genome, and has limitations in terms of resolving the complex and mixed metagenome sequences. These limitations can be overcome by a technology providing long-read sequences: PacBio RS II sequencing, based on single-molecule and real-time (SMRT) detection, dramatically improves the assembly of a metagenome from long-read sequences, on average 3,000 bp or longer than 20,000 bp, compared with the conventional next-generation sequencing technologies [12, 13]. One study showed that PacBio sequencing can provide high-quality information on lignocellulose-degrading bacteria and fungi via genes and canonical pathways involved in lignocellulose decomposition [14].

Most metagenomic studies have focused on screening for microbial enzymes for lignocellulose decomposition in extreme microenvironments: low pH, high moisture, or high temperature [9–11]. Industrial bagasse collection serves as a suitable ecological niche for studying microbial lignocellulose deconstruction owing to its physiologically amenable conditions [7]. Nevertheless, because the industrial bagasse collection is conducted after various types of extreme pretreatments (*e.g.*, high-temperature processing) that change the natural conditions, it may not accurately represent the natural microbial metabolic processes or allow us to discover microbial enzymes with decomposition activities under moderate conditions that do not require any external energy supply for biotechnological exploitation. To explore the lignocellulose-degrading enzymes in microbes, it may be one method for studying dead plant matter or its debris, which are more readily decomposed by microbes. Thus, it is necessary to study a microbial community or its functionality in plant dry matter exposed to natural conditions for obtaining appropriate genetic information on the degradation of biomass under moderate conditions.

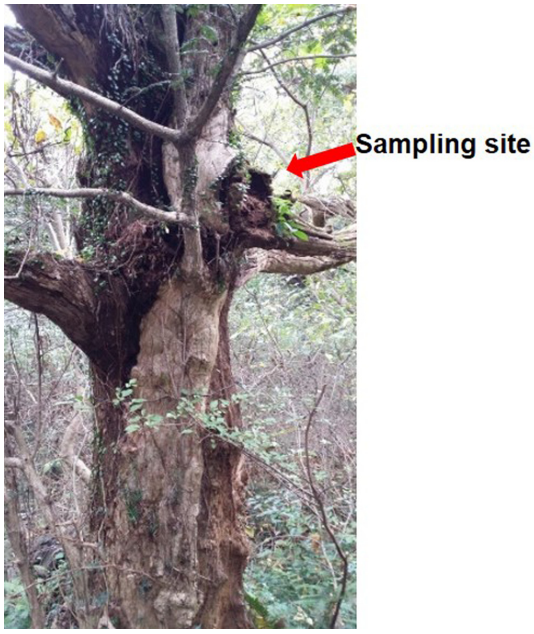

Fig. 1. Picture of the sampling site (33° 29′ N, 126° 48′ E) in Bija forest, Jeju Island.
Samples were collected from the detritus of rotten branches of *Torreya nucifera* (L).

Torreya nucifera (L) (Bija) on Jeju Island, Korea, is located in the Bija forest (33° 29′ N, 126° 48′ E) at altitudes of 280–380 m and the Bija forest is located in a region with a special geographical feature called Gotjawal (Fig. 1). *T. nucifera* (L) (Bija) is a coniferous tree, which contains 40–50% of cellulose, 25–30% of hemicellulose, and 25–35% of lignin, in general. The annual temperature of Bija forest ranges from 14°C to 21°C, the annual precipitation is 1,474.9 mm according to the Korea Meteorological Administration (2001–2010), and the forest has been reported as a valuable repository of biological diversity [15]. It is important to study the microbial decomposition of plant biomass and lignocellulose in Bija forest because it is located in Gotjawal, which is under moderate environmental conditions, and therefore it is predicted that lignocellulose degradation by microbes occurs naturally.

Here, we focused on the microbial metabolic capacity for degrading lignocellulose in wood detritus of *T. nucifera* (L) (Bija). We examined the following topics: (i) lignocellulose-degrading microbial community structure based on 16S rRNA gene sequencing (Illumina), (ii) the lignocellulose degradation process by a microbial enzyme detected in a

metagenome via the long-read SMRT sequencing technology, and (iii) new candidates for lignocellulose-degradative enzymes under moderate conditions without pretreatments. The characterization of the metagenome in Bija's wood detritus will provide basic knowledge on the ecological role of bacteria and their enzymatic activities for lignocellulosic biomass degradation.

Materials and Methods

Sampling Site and DNA Extraction

The wood detritus was collected from stems of nut-bearing *Torreya* (*nucifera* (L) Siebold & Zucc., Administration No. 0973) in the Korean natural monument Jeju Bija forest, South Korea in November 2011 (Fig. 1). A part of a decayed branch of a living *T. nucifera* (Bija) tree located at about 150 cm of height from the ground had been cut off, and the wood detritus was found on the residual proximal part of the branch. The detritus looked like soil and had dark brown color, all soft texture, and powder-like structure. A total of 20 g of the above wood detritus was gathered with a spatula and stored at 4°C until DNA extraction. Approximately 2.4 g of sample was used for metagenomic DNA extraction with the PowerSoil DNA Isolation Kit (MO BIO Laboratories Inc., USA).

PCR and Illumina MiSeq Sequencing

Extracted DNA was used for PCR and 16S rRNA gene sequencing. The PCR amplification was carried out using the primer set V4-V5 region of the 16S rRNA gene (forward: CCA GCA GC[T,C] GCG GT[G,A] A.; reverse: CCG TCA ATT C.T TT[G,A] AGT). In the first PCR, thermal cycle conditions were 95°C for 3 min, followed by 33 cycles of 30 sec denaturation at 95°C, 30 sec at 55°C for annealing, and 1 min at 72°C for extension, and a final extension was performed at 72°C for 5 min. After verifying that PCR amplification was successful using gel electrophoresis with Difco agar noble (Becton, Dickinson and Company, USA), the PCR product was pooled for cleanup using AMPure XP beads (Beckman Coulter, UK) and the product was incubated with 10 mM Tris buffer (pH 8.5). Then, the product was barcoded using the Nextera XT Index Kit v2 (Illumina, USA) and amplified for 8 cycles. The second PCR product was purified using AMPure XP beads and the quality of the enriched libraries was evaluated using the Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, US). The libraries were sequenced at Macrogen (Korea) by means of the Illumina MiSeq platform, and the generated 16S rRNA gene sequences were processed and analyzed with QIIME [16, 17].

The raw reads were merged using the Python scripts *iu-merge-pairs* [18]. The paired-end sequences were clustered into operational taxonomic units (OTUs) using UCLUST algorithm at a genetic divergence level of 3% [19]. For this analysis, the abundant sequence within an OTU was picked by the QIIME scripts *pick_otus.py* and *pick_rep_set.py*, and the RDP taxonomy with minimum confidence value of 0.5 was conducted according to the script *assign_taxonomy.py* [20]. Among the 16,283 OTUs, all OTUs identified as belonging to class "Chloroplast" (105 OTUs, 1,239 sequences) or "Mitochondria" (801 OTUs, 6,614 sequences), which comes from Eukaryotic cell, were removed using the scripts *filter_taxa_from_otu_tably.py* and *filter_fasta.py*. Before we inferred a phylogenetic tree relating the sequences, *align_seqs.py* aligned the sequences in a FASTA file to each other using PyNAST as an aligner [21]. The aligned sequences were filtered to generate a useful phylogenetic tree when aligning against a template using *filter_alignment.py* and a phylogenetic tree relating the OTUs from a multiple sequence alignment by *make_phylogeny.py*. An OTU table was created using *make_otu_table.py*, and the *summarize_taxa.py* script provided a summary of the relative abundance of the representation of taxonomic groups. The OTUs of the most abundant unclassified genera were matched to closest RDP sequences using online RDP tools, Sequence Match (SeqMatch) [20]. The 16S rRNA gene sequences in this paper have been deposited in the Sequence Read Archive (SRA) database of the NCBI (Accession No. SRR5229899).

Metagenome Sequencing

Three SMRT cells that each contained ~80 μg of extracted metagenomic DNA were run on the PacBio RS II system using P6 C4 chemistry (PacBio, USA). For de novo assembly, a hierarchical genome assembly process workflow (HGAP ver. 2.3) packaged in the SMRT analysis software was used [22]. During the assembly process, all the reads were assembled with 3−8-kb seed base size. The short assembled contigs were aligned against long continuous contigs for error correction. Owing to large numbers of assembled reads and large contig lengths, a seed library of 4 kb was selected. For bacterial genome annotation, Prokka (rapid prokaryotic genome annotation) was applied [23]. All protein-coding features (CDS) were found using Prodigal 2.6 and similarity searches against protein sequence libraries by BLAST+ [24, 25]. Among the total annotated 2,872 CDS, the analyzed ones were filtered with 300 bp in length. Taxonomic classification of contigs was performed by Edge Perl scripts [26] using Burrows-Wheeler Aligner (BWA) [27] mapping to National Center for Biotechnology Information (NCBI) RefSeq [26−28]. The data reported in this paper are available at the SRA database with the accession number of SRR5230003.

Analysis of Lignocellulose-Degrading Enzymes

Amino acid sequences translated from Open Reading Frames (ORFs) were generated by Prokka and were continuously searched to obtain a candidate lignocellulose-degrading enzyme in a carbohydrate-active enzyme database (CAZy database; http://www.cazy.org/) [29, 30]. A carbohydrate-active enzyme analysis toolkit (CAT; http://mothra.ornl.gov/cgi-bin/cat/cat.cgi/) [31] was also employed, and enzymes were filtered with an e-value of 0.01. For the other functional profiles, amino acids sequences was used for searches in Pfam: the protein families database (Pfam-A, ver. 30) [32], which is defined by hidden Markov model (HMM) profiles. The same HMM was used to search against ORFs

translated from the contigs of the metagenomes using hmmsearch 3.1. The results satisfying a bit score of 25 or more and e-value of 0.01 were selected. The additional annotation for the contig was done by the RAST (Rapid Annotation using Subsystem Technology) server [33]. The overall functional features of metagenome data including cluster of orthologous groups (COG) and Kyoto Encyclopedia of Genes and Genomes (KEGG) orthology were also obtained by means of Web-based eggNOG and GhostKOALA tools, respectively [34, 35]. In the analysis of COGs, the mapping mode was HMMER based on the all-Bacteria database.

Results

Composition of the Population of Bija's Wood Detritus-Degrading Bacteria

We performed 16S rRNA gene amplicon sequencing using the Illumina MiSeq Platform to identify the taxonomic composition of the bacterial community degrading lignocellulosic biomass in *T. nucifera* (L) located in Bija forest. As results of the sequencing, a total of 366,621 pair-end reads were merged into 94,290 sequences. From all the 16S rRNA sequences, 16,283 OTUs were obtained, and 15,277 OTUs were obtained finally after removing the ones from Chloroplast and Mitochondria. A total of 99.9% of OTUs represented the kingdom Bacteria, 0.003% corresponded to Archaea, and the remaining were unclassified. Three major phyla were Proteobacteria (48.2%), Bacteroidetes (11.7%), and Actinobacteria (11.4%); members of these phyla collectively represented over 70% of all 16S rRNA sequences. The most abundant orders (>5% of total abundance) were Xanthomonadales (15.5%), Rhodospirillales (9.0%), Rhizobiales (7.1%), Saprospirales (5.9%), Gemmatales (5.4%), and Sphingobacteriales (5.3%); they collectively represented 48% of all 16S rRNA sequences. The most abundant genus was unclassified Xanthomonadaceae, followed by unclassified Sinobacteraceae and unclassified Chitinophagaceae. OTUs assigned to unclassified Xanthomonadaceae (most closely related to genus *Mizugakiibacter* and *Povalibacter* by SeqMatch) were the most dominant (7.2%) (Fig. 2).

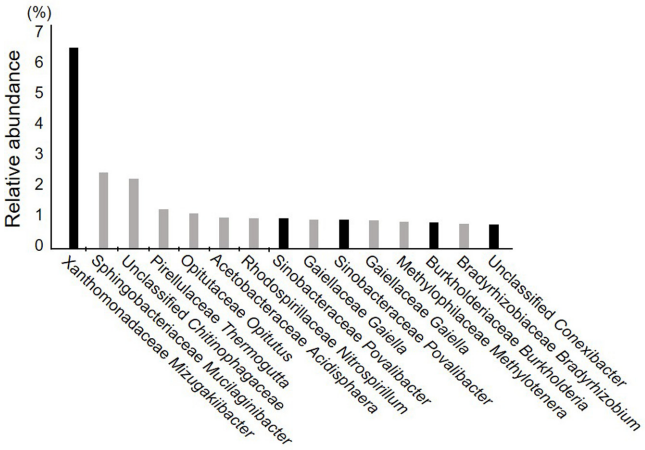

Fig. 2. Rank abundance curve for the most abundant OTUs.
Each bar indicates representative OTUs assigned to a taxon at the genus level. The black bars (Xanthomonadaceae *Mizugakiibacter*, Sinobacteraceae *Povalibacter*, and Unclassified *Conexibacter*) were assigned to Xanthomonadales, Burkholderiales, and Solirubrobacterales, and these three orders played an important role in lignocellulose degradation in Bija wood detritus.

Metagenomic Analysis for the PacBio Platform

We obtained metagenomic data from three SMRT cells, PacBio RS II in Bija's wood detritus. A total of 217,558 reads ranging from 5,202 to 61,553 bp in length were obtained and subsequently assembled by the hierarchical genome-assembly process (HGAP) into 204 contigs with a contig N50 of 15,225 bp. The mean coverage of the assembled sequences was 45.5-fold, and this coverage was relatively low, but accurate analysis was possible by providing the long reads to reduce the error through assembly. The metagenome dataset statistics are summarized in Table 1. The contigs have been classified using BWA mapping to NCBI RefSeq. Only two phyla, Proteobacteria (57.8%) and Actinobacteria (10.3%), were observed, and the most abundant orders were Xanthomonadales (39.2%), Burkholderiales (7.8%), and Solirubrobacterales (4.4%) (Table 2).

We compared results on bacterial community composition obtained from the 16S rRNA genes from Illumina and taxonomic information of contigs from PacBio metagenome by BWA mapping. Both results revealed the dominance of Proteobacteria and Actinobacteria in Bija's wood detritus, and two orders (Xanthomonadales and Rhizobiales) dominated. We found that both technologies yield some similar dominant members of microbial community composition, although their relative proportions varied between the two technologies. For example, order Xanthomonadales had a

Table 1. Summary of the metagenome using PacBio® RS II.

Parameters	Sequencing results
Number of bases (bp)	3,464,830,886
Number of reads (bp)	226,667
Number of contigs	204
Mean coverage (×)	45.45
N50 contig length (bp)	15,225
Max contig length (bp)	61,553
Number of ORFs	2,480

Table 2. Summary of metagenomic contigs compared with ORFs containing CAZy families.

Phylum	Order	Number of metagenome contigs*	Number of ORFs containing CAZymes	Number of module ORFs
Proteobacteria (57.8%)	Aeromonadales	1 (6)	1	1
	Burkholderiales	16 (218)	33	5
	Chromatiales	1 (8)	6	1
	Enterobacteriales	2 (34)	3	-
	Myxococcales	2 (18)	2	-
	Neisseriales	1 (19)	1	-
	Oceanospirillales	1 (15)	3	-
	Pseudomonadales	2 (51)	6	3
	Rhizobiales	7 (48)	8	-
	Rhodobacterales	1 (6)	1	-
	Rhodocyclales	2 (32)	9	2
	Rhodospirillales	2 (22)	1	-
	Xanthomonadales	80 (1,208)	199	35
Actinobacteria (10.3%)	Corynebacteriales	2 (18)	2	-
	Frankiales	1 (13)	1	-
	Geodermatophilales	1 (24)	2	1
	Kineosporiales	1 (7)	-	-
	Micrococcales	1 (7)	1	-
	Micromonosporales	1 (9)	1	-
	Propionibacteriales	2 (16)	5	1
	Solirubrobacterales	9 (95)	13	5
	Streptomycetales	3 (41)	7	3
Unclassified phyla		65 (566)	102	8
Total		204 (2,480)	403	65

*The values represent the number of ORFs assigned to each taxonomy.

low relative abundance rates (15.5%) in 16S rRNA gene analysis, but it appeared at 39.2% in the metagenomics data, and order Rhizobiales was accounted for as 7.1% in 16S rRNA gene analysis whereas it had a relative abundance of 3% in the metagenomics data (Table 2). Two orders showed large distribution in both results commonly; we investigated 16S rRNA gene analysis at the genus level. It was supposed that the most abundant genus was *Mucilaginibacter* of Xanthomonadales (RDP SeqMatch results), which was also found in the OTU analysis the most frequently, and the second largest genus of Xanthomonadales was *Povalibacter*. Among the Burkholderiales members, the most abundant genus was considered to be *Burkholderia*, which accounted for 0.9% in the OTU analysis, and the greatest number of Solirubrobacterales was *Conexibacter*.

ORFs were predicted and annotated for each contig by means of the Prodigal and rapid prokaryotic genome annotation software (Prokka) [23, 25], and 2,480 ORFs were identified. All the predicted ORFs were assigned to COG for functional profiling and were also used for analyses in KEGG for metabolic pathway identification. Among all the ORFs, 2,049 were assigned to a specific COG belonging to one of the following functional categories: "metabolism," "information storage and processing," and "cellular process and signaling." Carbohydrate transport and metabolism, which were the most relevant to lignocellulose degradation, represented 14.5% of all the categories of metabolism. Amino acid transport and metabolism, energy production and conversion, and carbohydrate transport and metabolism represented 53% of the other classes of metabolism.

Metabolic-profile mapping to the KEGG database showed results similar to those of COG assignments. The predominant pathways belonged to metabolism (51%), genetic information processing (20%), and environmental information processing (13%). The subcategories of metabolism were amino acid metabolism (18.2%), carbohydrate metabolism (15.3%),

metabolism of cofactors and vitamins (10.8%), and energy metabolism (8.7%). Detailed analysis of the subcategory "carbohydrate metabolism" enabled detection of a microbial enzyme related to lignocellulosic biomass degradation, with fructose/mannose (6.6%) and galactose (3.8%) as the intermediate products of lignocellulose degradation.

The Lignocellulose-Degrading Metabolic Potential

To gain insights into the lignocellulosic degradation abilities of the microbial community in Bija's wood detritus, the predicted ORFs were mapped to the CAZy database using Prokka [29, 30]. The 403 ORFs from the total of 2,480 annotated ORFs were assigned to five carbohydrate-active enzyme families from six enzyme families in the CAZy database. The enzymes obtained from Bija's wood detritus were assigned to the glycosyltransferase family (GT; 154 ORFs or 38.2% of total CAZymes), glycoside hydrolase family (GH; 152 or 37.7%), carbohydrate-binding module family (CBM; 42 or 10.4%), auxiliary activity family (AA; 28 or 6.9%), and carbohydrate esterase family (CE; 27 or 6.7%). Of these, 121 (30%) CAZymes possess activity related to the degradation of lignocellulosic substrate, including 19 different families of GHs. Amongst the cellulases, three families (GH5, 9, and 12), which include endoglucanases, exoglucanases, and cellobiohydrolases, were found in Bija's detritus (Table 3). For hemicellulose degradation, 14 families including GH10, 16, 76, and 53 were present, and GH53, endo-1,4-β-galactanase, was the most abundant hemicellulase, followed by GH92 (alpha-mannosidase), in the metagenome of Bija's detritus. The CAZymes encoding endo-1,4-β-galactanase (GH53) might reflect the requirement to decompose the greater abundance of galactan as compared with xylan and mannan. The downstream decomposition of hemicellulose was mainly catalyzed by various oligosaccharide-degrading enzymes such as β-glucosidase (GH3) and β-xylosidase (GH39). We also found six families of AAs known as lignin degradation enzymes, and the majority of lignin-degrading enzymes were GMC oxidoreductase (GMC: glucose-methanol-choline; AA3, 43% of all AAs). The multicopper oxidase (AA1) and glucooliggosachharide oxidase (AA7) were also observed, followed by AA4, 5, and 6. In addition, a number of predicted ORFs were mapped to carbohydrate-binding modules that adhered to substrates, supported other enzymes, and were commonly associated with several GHs to help the GHs with binding to target substrates. CBM35 participates in lignocellulose degradation by binding to xylan, and CBM32 was shown to be a galactose-binding enzyme.

We compared CAZymes with results from the Pfam database to identify the consensus plant biomass degradation modules, from M1 to M5 [36]. They assigned specific CAZymes and protein families to five consensus degradation modules. Modules 1, 2, and 3 represent lignocellulose degradation modules, xylan binding and xyloglucan degradation modules, and pectin degradation modules, respectively. M4 corresponds to degradation of glycan compounds, and M5 is defined as a structural component of the cellulosome-based degradation paradigm. A total of 65 proteins were found for these five modules (Table 2). Our results indicated that some GHs (GH5, 9, and 10) and

Table 3. Distribution of CAZymes according to lignocellulose degradation substrates identified from the metagenome.

CAZy families	Known activities	Count
Lignin		
AA1	Multicopper oxidase	4
AA3	GMC oxidoreductase	12
AA4	Vanillyl alcohol oxidase	2
AA5	Radical-copper oxidase	3
AA6	1,4-Benzoquinone reductase	3
AA7	Glucooligosaccharide oxidase	4
Cellulose		
GH5	Cellulase (exo-/endoglucanase)	1
GH9	Endoglucanase/cellobiohydrolase	2
GH12	Endoglucanase and xyloglucan hydrolysis	2
Hemicellulose		
GH10	Endo-1,4-β-xylanase	3
GH12	Endoglucanase and xyloglucan hydrolysis	2
GH16	Xyloglucanases:xyloglycosyltransferases	12
GH1	β-Glucosidase and other β-linked dimers	4
GH3	Mainly β-glucosidases	9
GH39	β-Xylosidase	5
GH76	α-1,6-Mannase	8
GH38	α-Mannosidase	1
GH31	α-Glucosidase	2
GH92	α-Mannosidase	12
GH53	Endo-1,4-β-galactanase	21
GH32	α-Galactosidase	2
GH35	β-Galactosidase	2
GH27	α-Galactosidase	1
Other oligosaccharide-degrading enzymes		
GH15	α-Trehalose	2
GH2	β-Galactosidases and other β-linked dimers	3
Total AAs/% of total CAZymes		28 (6.9%)
Total GHs/% of total CAZymes		156 (38.7%)

CBM35 were included in M1, consensus lignocellulose degradation modules, whereas GH16 and some CBMs (CBM32, 47, and 13) were affiliated with M2, which are xylan-binding and xyloglucan degradation modules. In addition, PF00759, PF00331, and PF13472 were assigned to M1, whereas PF13229, PF13472, and PF13524 were assigned to the M3 module (pectin degradation modules).

Lignocellulosic Biomass-Degrading Bacteria

To investigate the taxonomic origins of the lignocellulose degradation enzymes, we mapped all the annotated ORFs to the taxon database, which is a result of BWA mapping to NCBI RefSeq. We compared the total number of metagenome contigs with CAZymes having lignocellulose-degrading capacity in ORFs at the order level (Fig. 3). It was confirmed that CAZymes related to lignocellulose degradation are proportional to the number of total metagenomes. Xanthomonadales (Gammaproteobacteria), Burkholderiales (Betaproteobacteria), and Solirubrobacterales (Actinobacteria) were widely distributed in the Bija metagenome, and many of the lignocellulose-degrading enzymes were assigned to the ORFs of these orders. We observed the correlation of the microbes capable of degrading lignocellulose and constituents of lignocellulosic biomass. Overall, a higher abundance of members of Proteobacteria was detected, and among them, Xanthomonadales was the largest bacterial order, which had remarkable abundance in entire lignocellulose degradation in Bija's detritus (Table 4). Among the genes affiliated to the plant biomass degradation module of lignocellulose degradation, 70.8% of enzymes were located in the contigs belonging to three orders: Xanthomonadales, Burkholderiales, and Solirubrobacterales. In particular, the Xanthomonadales

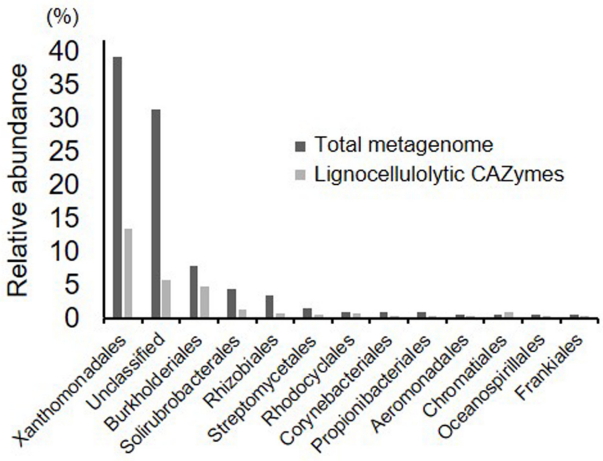

Fig. 3. Bar graph of relative abundance of lignocellulolytic CAZymes and contigs of the most abundant orders in the metagenome.

have at least one enzyme in each of five modules, and many enzymes involved in the decomposition of each constituent of lignocellulose were found in ORFs of Xanthomonadales.

Focusing on GH and AA that are directly related to lignocellulose degradation, the CAZymes related to degradation of lignin were assigned to the orders Burkholderiales and Xanthomonadales (Fig. 4). In Bija forest, once cellulose is degraded by the Xanthomonadales, cellobiose is broken down by the Rhizobiales and Burkholderiales. Xylan, known to account for the largest proportion of hemicellulose, was degraded by various bacterial orders of Actinobacteria and Proteobacteria. Xanthomonadales were found to be involved in the degradation of mannan, and the mannan dimer was

Table 4. Consensus plant degradation modules from the metagenome.

Order	M1 (Lignocellulose)	M2 (Xylan, xyloglucan)	M3 (Pectin)	M4 (Glycan compounds)	M5 (Cellulosome based)
Aeromonadales				1	
Burkholderiales				5	
Chromatiales				1	
Geodermatophilales		1			
Propionibacteriales	1				
Pseudomonadales	1	2			
Rhodocyclales	1			1	
Solirubrobacterales		2	2	1	
Streptomycetales		1		2	
Xanthomonadales	7	10	3	14	1
Unclassified	2	4			2
	11	21	5	25	3

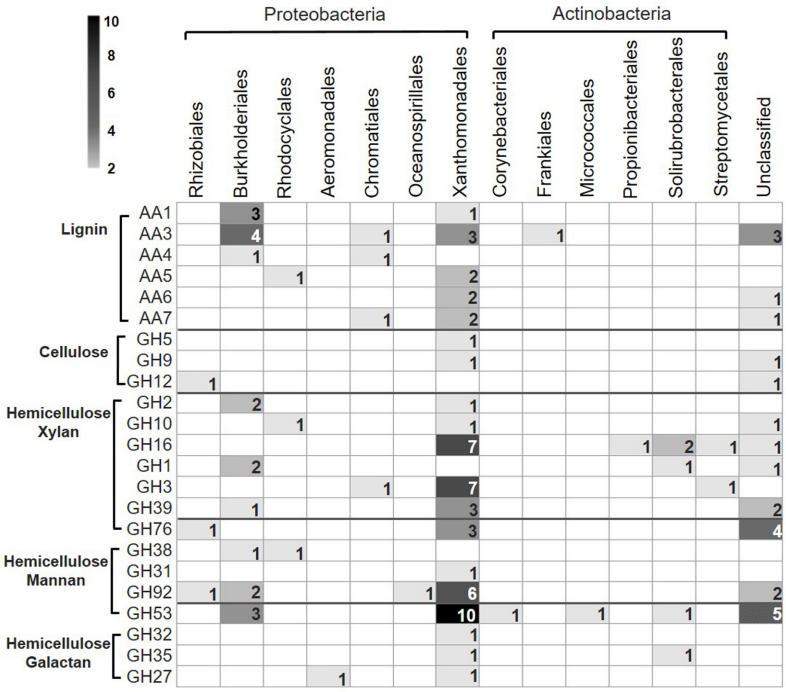

Fig. 4. Distribution of lignocellulolytic CAZymes (on left) with the most abundant orders (on the top) found in the metagenome. Gray shades and the numbers in each cell represent the counts of CAZymes found in each bacterial order.

degraded by other members of Proteobacteria. The enzymes that break down galactan were found in Xanthomonadales, and galactan dimer-degrading enzymes were produced from Actinobacteria as much as Proteobacteria. Overall, Proteobacteria dominated the lignocellulose degradation of Bija's wood detritus, and members of Actinobacteria were involved only in hemicellulose degradation. It was supposed that the different metabolic activities depending on microbes had a preference for specific substrates. In the metagenome of Bija forest, the predominant orders were dominant in the degradation of lignocellulose and a continuous appearance of lignocellulolytic enzymes was found in contigs of these orders. For example, enzymes AA3 (GMC oxidoreductase), AA4 (vanillyl alcohol oxidase), and AA7 (glucooligosaccharide oxidase) were found in the same contig of Chromatiales, which played a role in lignin degradation, and we found that the contig was assigned to the complete genome of *Halorhodospira halophila* SL1, using RAST [33]. Glycolate dehydrogenase, which was known to modify the lignin and aromatic compounds, was found in the genome, and the genes were matched to AA4 and AA7 [37, 38].

Discussion

T. nucifera (L) (Bija) is a coniferous tree growing in specific subtropical regions, such as Jeju Island in Korea. Its detritus is associated with microorganisms and is an important ecosystem component for studying microbial ecology of natural lignocellulose decomposition. As well as hardwood, a softwood such as *T. nucifera* (Bija) can be the material, and a proportion of lignocellulose components is different so that the requirement for production of ethanol is different according to the materials [39]. Virtually nothing is known about this activity or the relative importance of bacteria in the decomposition of Bija's detritus under mild conditions. Because our study was the first to estimate the taxonomic and functional structures of bacteria participating in lignocellulose decomposition of Bija's detritus, a direct comparison with other studies on wood detritus cannot be made. Nonetheless, 16S rRNA gene and metagenome analyses allowed us to gain insights into the bacterial activities and processes involved in lignocellulose decomposition.

We analyzed metagenomic data using PacBio RS II to study the microbial community and to determine functional profiles of the metagenome. The SMRT sequencing technology (PacBio) has an advantage, in that long reads can be obtained, and highly accurate consensus sequences are generated [13]. A total of 2,883,759 bases were generated with high coverage (45.45), which provided sufficient depth to analyze the gene syntax in each scaffold. The long

contigs of our assembled metagenome sequences from PacBio sequencing produced 2,478 coding sequences that were greater than 300 bp in length and allowed us to reveal the raw diverse microbial community from ~70% of contigs.

The main findings of this study confirmed that the dominant members—Burkholderiales (Proteobacteria), Xanthomonadales (Proteobacteria), and Solirubrobacterales (Actinobacteria)—play an important role in lignocellulose decomposition, as shown in other studies [40, 41]. At the genus level, *Xanthomonas*, *Psuedomonas*, and *Stenotrophomonas* appeared in corn stover and played a role in degradation of cellulose and hemicellulose [42]. *Burkholderia* species (*sacchari*, *tuberum*, *etc*.) degrade cellulose in beech wood from forest soil, and *Conexibacter* in Solirubrobacterales has potential for degrading cellulose in soil from Caatinga [43, 44]. These results suggested that relatively few bacterial groups comprise the majority in various microenvironments conducive to lignocellulose degradation, although the proportion of dominant groups varies among the environment types [8–11]. The low diversity at phylum or order levels is likely created by the limitedness of nutrients from lignocellulosic components.

In our study, a few lignocellulose-degrading bacteria such as Xanthomonadales (Gammaproteobacteria) have lignocellulolytic enzymes based on five consensus plant biomass degradation modules, and have relations between loci in genomes and functions associated with degradation. Proteobacteria are abundant in sugarcane bagasse, where biomass is actively degraded by the lignocellulolytic microbiota and has various endo-acting enzymes degrading cellulose, hemicellulose, and lignin [7]. Xanthomonadales, the most abundant order of Proteobacteria in Bija's metagenome, was studied as the soil microbes that are metabolically active on an oxic surface of paddy soil and decompose plant materials in cooperation with fungi [45]. The next largest order of phylum Proteobacteria is Burkholderiales. *Burkholderia* species are capable of degrading plant materials comprising cellulose, hemicellulose, lignin, and xylose; genes encoding lignin-degrading enzymes such as catalases and peroxidases were found among *Burkholderia* species genes [14].

The results of COG and KEGG mapping showed that metabolism, especially carbohydrate metabolism, which is considered a lignocellulosic biomass degradation-related category, was found relatively more frequently than other functional categories. This finding suggests that the microbial community in Bija's wood detritus is metabolically active from the standpoint of lignocellulose decomposition. The analysis of microbial enzyme families enabled us to identify some CAZymes participating in lignocellulose degradation. Microbial communities that have a potential for lignocellulosic biomass degradation vary according to the various environments, and microbial enzymes related to direct decomposition of the biomass are different as well [46]. In wood detritus, the phylum that has abundant carbohydrate-active enzymes is Proteobacteria, and the distributions of CAZymes of the specific phylum are different depending on environments such as rice straw-adapted microbial consortia (Actinobacteria) and the termite hindgut (Fibrobacteres and Spirochetes) [47, 48]. Our study identified Burkholderiales as apparently responsible for lignin degradation, and Xanthomonadales with a capacity for degrading overall lignocellulosic substrates, including mainly hemicellulose. As for the presence of protein families, each order has different enzymes related to lignocellulose degradation, and this result indicates that the degradation capacity of microorganisms is different according to the specific substrates and constituents of lignocellulose. Although many GH16 and GH53 proteins (hemicellulolytic enzymes) were found, the genes belonging to families GH30 and GH43 were not found in the metagenome of wood detritus. Moreover, these two enzymes, which were not found in the metagenome, were detected in the forest floor genome [49]. Thus, these results suggest that the enzymes used to degrade lignocellulose are different for each environment.

Because the industrial degradation of lignocellulose to valuable products like ethanol is costly and requires harsh conditions such as high temperature and an acid environment, the study of microbial genomic potential for lignocellulose degradation under natural conditions should contribute to the development of eco-friendly decomposition of lignocellulosic biomass. Bija forests have a unique environment and biological diversity, and we found that the microbes in the Bija forest under study have an ability to decompose lignocellulose. The combined results gathered from the metagenome of detritus revealed that various microbes capable of degrading lignocellulose have different genes and thus are selective for degradation. It is expected that the understanding of natural microbial enzymatic processes in Bija forests may result in useful methods for industrial applications without the high cost and harsh conditions.

Acknowledgments

This research was supported by a Chung-Ang University research scholarship grant in 2014, and by Basic Science

Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2014R1A1A2A16055779).

References

1. Lynd LR, Weimer PJ, Van Zyl WH, Pretorius IS. 2002. Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol. Mol. Biol. Rev.* **66:** 506-577.
2. Lee H, Hamid S, Zain S. 2014. Conversion of lignocellulosic biomass to nanocellulose: structure and chemical process. *ScientificWorldJournal.* **2014:** 631013.
3. Himmel ME, Ding S-Y, Johnson DK, Adney WS, Nimlos MR, Brady JW, *et al*. 2007. Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* **315:** 804-807.
4. Gallezot P. 2012. Conversion of biomass to selected chemical products. *Chem. Soc. Rev.* **41:** 1538-1558.
5. Kato DM, Elia N, Flythe M, Lynn BC. 2014. Pretreatment of lignocellulosic biomass using Fenton chemistry. *Bioresour. Technol.* **162:** 273-278.
6. Iqbal HMN, Kyazze G, Keshavarz T. 2013. Advances in the valorization of lignocellulosic materials by biotechnology: an overview. *BioResources* **8:** 3157-3176.
7. Mhuantong W, Charoensawan V, Kanokratana P, Tangphatsornruang S, Champreda V. 2015. Comparative analysis of sugarcane bagasse metagenome reveals unique and conserved biomass-degrading enzymes among lignocellulolytic microbial communities. *Biotechnol. Biofuels* **8:** 16.
8. Kanokratana P, Uengwetwanit T, Rattanachomsri U, Bunterngsook B, Nimchua T, Tangphatsornruang S, *et al.* 2011. Insights into the phylogeny and metabolic potential of a primary tropical peat swamp forest microbial community by metagenomic analysis. *Microb. Ecol.* **61:** 518-528.
9. Woo HL, Hazen TC, Simmons BA, DeAngelis KM. 2014. Enzyme activities of aerobic lignocellulolytic bacteria isolated from wet tropical forest soils. *Syst. Appl. Microbiol.* **37:** 60-67.
10. Aylward FO, Burnum KE, Scott JJ, Suen G, Tringe SG, Adams SM, *et al.* 2012. Metagenomic and metaproteomic insights into bacterial communities in leaf-cutter ant fungus gardens. *ISME J.* **6:** 1688-1701.
11. Scully ED, Geib SM, Hoover K, Tien M, Tringe SG, Barry KW, *et al.* 2013. Metagenomic profiling reveals lignocellulose degrading system in a microbial community associated with a wood-feeding beetle. *PLoS One* **8:** e73827.
12. Metzker ML. 2010. Sequencing technologies — the next generation. *Nat. Rev. Genetics* **11:** 31-46.
13. Roberts RJ, Carneiro MO, Schatz MC. 2013. The advantages of SMRT sequencing. *Genome Biol.* **14:** 405.
14. Qin W. 2016. Recent developments in using advanced sequencing technologies for the genomic studies of lignin and cellulose degrading microorganisms. *Int. J. Biol. Sci.* **12:** 156.
15. Kim DS, Lee JH, Yang SH. 2010. *Plant Community Dynamics*, pp. 107-135.
16. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, *et al.* 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6:** 1621-1624.
17. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, *et al.* 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7:** 335-336.
18. Eren AM, Vineis JH, Morrison HG, Sogin ML. 2013. A filtering method to generate high quality short reads using Illumina paired-end technology. *PLoS One* **8:** e66643.
19. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26:** 2460-2461.
20. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, *et al.* 2009. The ribosomal database project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* **37:** D141-D145.
21. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. 2010. PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26:** 266-267.
22. Sakai H, Naito K, Ogiso-Tanaka E, Takahashi Y, Iseki K, Muto C, *et al.* 2015. The power of single molecule real-time sequencing technology in the de novo assembly of a eukaryotic genome. *Sci. Rep.* **5:** 16780.
23. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30:** 2068-2069.
24. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11:** 119.
25. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, *et al*. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10:** 421.
26. Li P-E, Lo C-C, Anderson JJ, Davenport KW, Bishop-Lilly KA, Xu Y, *et al.* 2017. Enabling the democratization of the genomics revolution with a fully integrated Web-based bioinformatics platform. *Nucleic Acids Res.* **45:** 67-80.
27. Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26:** 589-595.
28. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, *et al.* 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44:** D733-D745.
29. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The carbohydrate-active enzymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* **37:** D233-D238.
30. Lombard V, Ramulu HG, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42:** D490-D495.
31. Park BH, Karpinets TV, Syed MH, Leuze MR, Uberbacher EC.

2010. CAZymes Analysis Toolkit (CAT): Web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database. *Glycobiology* **20:** 1574-1584.
32. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, *et al.* 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44:** D279-D285.
33. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, *et al.* 2008. The RAST server: rapid annotations using subsystems technology. *BMC Genomics* **9:** 75.
34. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, *et al.* 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* **44:** D286-D293.
35. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **428:** 726-731.
36. Konietzny SG, Pope PB, Weimann A, McHardy AC. 2014. Inference of phenotype-defining functional modules of protein families for microbial plant biomass degraders. *Biotechnol. Biofuels* **7:** 124.
37. Zhu D, Zhang P, Xie C, Zhang W, Sun J, Qian WJ, Yang B. 2017. Biodegradation of alkaline lignin by *Bacillus ligniniphilus* L1. *Biotechnol. Biofuels* **10:** 44.
38. Zhang J, Presley GN, Hammel KE, Ryu JS, Menke JR, Figueroa M, *et al.* 2016. Localizing gene regulation reveals a staggered wood decay mechanism for the brown rot fungus *Postia placenta*. *Proc. Natl. Acad. Sci. USA* **113:** 10968-10973.
39. Horn SJ, Vaaje-Kolstad G, Westereng B, Eijsink VG. 2012. Novel enzymes for the degradation of cellulose. *Biotechnol. Biofuels* **5:** 45.
40. Kameshwar AKS, Qin WS. 2016. Recent developments in using advanced sequencing technologies for the genomic studies of lignin and cellulose degrading microorganisms. *Int. J. Biol. Sci.* **12:** 156-171.
41. Han S-I. 2016. Phylogenetic characteristics of bacterial populations and isolation of aromatic compounds utilizing bacteria from humus layer of oak forest. *Korean J. Microbiol.* **52:** 175-182.
42. Jimenez DJ, de Lima Brossi MJ, Schuckel J, Kracun SK, Willats WG, van Elsas JD. 2016. Characterization of three plant biomass-degrading microbial consortia by metagenomics- and metasecretomics-based approaches. *Appl. Microbiol. Biotechnol.* **100:** 10463-10477.
43. Folman LB, Klein Gunnewiek PJ, Boddy L, de Boer W. 2008. Impact of white-rot fungi on numbers and community composition of bacteria colonizing beech wood from forest soil. *FEMS Microbiol. Ecol.* **63:** 181-191.
44. Lacerda Junior GV, Noronha MF, de Sousa ST, Cabral L, Domingos DF, Saber ML, *et al.* 2017. Potential of semiarid soil from Caatinga biome as a novel source for mining lignocellulose-degrading enzymes. *FEMS Microbiol. Ecol.* **93:** fiw248.
45. Kim Y, Liesack W. 2015. Differential assemblage of functional units in paddy soil microbiomes. *PLoS One* **10:** e0122221.
46. Cragg SM, Beckham GT, Bruce NC, Bugg TD, Distel DL, Dupree P, *et al.* 2015. Lignocellulose degradation mechanisms across the Tree of Life. *Curr. Opin. Chem. Biol.* **29:** 108-119.
47. Wang C, Dong D, Wang H, Muller K, Qin Y, Wang H, *et al.* 2016. Metagenomic analysis of microbial consortia enriched from compost: new insights into the role of Actinobacteria in lignocellulose decomposition. *Biotechnol. Biofuels* **9:** 22.
48. Warnecke F, Luginbühl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, *et al.* 2007. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450:** 560-565.
49. Lopez-Mondejar R, Zuhlke D, Becher D, Riedel K, Baldrian P. 2016. Cellulose and hemicellulose decomposition by forest soil bacteria proceeds by the action of structurally variable enzymatic systems. *Sci. Rep.* **6:** 25279.