Ecological functions and environmental fate of exopolymers of Acidobacteria

Costa, O.Y.A.

Chapter 6

Genetic potential of microbial communities involved in the degradation of a complex acidobacterial extracellular polymer

Ohana Y. A. Costa, Mattias de Hollander, Agata Pijl, Eiko E. Kuramae

Modified version published as: Costa OYA, De Hollander M, Pijl A, Liu BB, Kuramae EE (2020). Cultivation-independent and cultivation-dependent metagenomes reveal genetic and enzymatic potential of microbial community involved in the degradation of a complex microbial polymer. **Microbiome**, 8(76).

Abstract

Polysaccharides are the main components of extracellular polymeric substances (EPS), biopolymers synthesized by a wide range of strains of microorganisms. EPS are the constituents that preserve the tridimensional structure of biofilms, maintaining internal cohesion and promoting adhesion to surfaces. The elimination of biofilms is important for human health because those structures cause problems in hospitals and in food processing industries. To identify EPS degrading microorganisms, we performed Stable Isotope Probing (SIP) combined with metagenomics on topsoil litter amended with EPS of *Granulicella* sp. WH15 (WH15EPS). In addition, we coupled solid culture medium with metagenomics to detect and cultivate potential GH producers. Among all carbohydrate-active enzymes (CAZymes) detected, the most abundant families belonged to Glycoside Transferase (GT) families. Among the glycoside hydrolases (GH), the most abundant family in the metagenomics datasets was amylase family GH13. In the "heavy" fraction of the metagenomics SIP dataset, GH109 (α-N-acetylgalactosaminidases), GH117 (agarases,), GH50 (agarases), GH32 (invertases and inulinases), GH17 (endoglucanases), GH71 (Mutanases) families were more abundant in comparison with the controls. Those GH familes originated from microorganisms that are believed to be able to degrade WH15EPS and potentially applicable for biofilm deconstruction. Subsequent assembly of 4 metagenome-assembled genomes (MAGs) (unclassified *Proteobacteria*) also contained GH families of interest, involving mannosidases, lysozymes, galactosidases and chitinases. We demonstrated that functional diversity induced by the presence of WH15EPS in both culture-dependent and culture independent approaches was enriched in GHs, such as amylases and endoglucanases that could be applied in chemical, pharmaceutical and food industrial sectors. Furthermore, WH15EPS may be used for the investigation and isolation of yet unknown taxa, such as unclassified *Proteobacteria* and *Planctomycetes*, increasing the number of current cultured bacterial representatives.

Keywords: *Acidobacteria*, EPS, Metagenonics, Stable Isotope Probing, *Planctomycetes*, Carbohydrates

1. Introduction

Exopolysaccharides are the main and most studied components of extracellular polymeric substances (EPS), biopolymers synthesized by a wide range of strains of microorganisms (Flemming & Wingender, 2010, Costa *et al.*, 2018). EPS are the constituents that preserve the tridimensional structure of biofilms, maintaining internal cohesion and promoting adhesion to surfaces (Flemming & Wingender, 2010). The elimination of biofilms is important for human health in general, because those structures are implicated in several diseases, causing problems for instance in hospitals and in food processing industries (Hunter, 2008, Nahar *et al.*, 2018). Enzymatic removal of biofilms is superior to the use of conventional cleaning agents, which are not eco-friendly, producing toxic residues and erosion of equipment (Nahar *et al.*, 2018). Enzymes are an environmentally friendly alternative due to their biodegradable nature (Liu & Kokare, 2017). EPS and biofilms are complex, requiring a wide range of enzymes for a complete degradation (Flemming & Wingender, 2010).

Glycoside hydrolases (GHs) are hydrolytic enzymes which can be applied for the degradation of EPS polysaccharides for the removal of biofilms (Fleming *et al.*, 2016). Furthermore, GHs are among the industrially important enzymes that are extensively searched through metagenomics, as they are extremely desired and important in food and other industrial sectors. Those enzymes are employed for brewing, baking, production of syrups, food processing, texture, flavoring, as well as the production of dairy and fermented foods (Coughlan *et al.*, 2015). GHs are also necessary for the production of biofuels, by converting cellulose and lignocellulosic biomass into sugars that can be fermented into bioethanol by microorganisms (Ezeilo *et al.*, 2017). More than 50% of the current industrial enzymes are produced by microorganisms, such as strains of *Bacillus* and *Aspergillus*, while around 15% are derived from plants (Liu & Kokare, 2017). In addition, microbial enzymes with potential applications were obtained from habitats such as hydrothermal vents (Legin *et al.*, 1997), arctic tundra (Oh *et al.*, 2019), cow rumen (Hess *et al.*, 2011) and termite guts (Warnecke *et al.*, 2007).

The main goal of our study was to use a microbial EPS to find microbes and functions involved in EPS degradation. We used topsoil-plant litter as a source of microbes. Plant litter is mostly composed of recalcitrant biopolymers, which are sources of carbon, energy and nutrients for microbial communities living in litters layers (Urbanová *et al.*, 2015, Vivanco *et al.*, 2018). Due to their complexity, the breakdown of plant cell wall components requires a wide range of enzymes, produced by the microorganisms during litter decomposition process (Schneider *et al.*, 2012, Chen *et al.*, 2018). Therefore, it is an interesting environment for the retrieval of complex polysaccharide-degrading enzymes. We applied the EPS of the *Acidobacteria Granulicella* sp. strain WH15 (WH15EPS). The unique composition of its EPS may be a source to retrieval of a wide range of novel glycoside hydrolase genes that could be employed in the industry for several processes. WH15EPS has a more complex composition than most commercially available microbial polymers. It is composed of 7 monosaccharides

(mannose, glucose, galactose, xylose, rhamnose, glucuronic and galacturonic acids) (Kielak *et al.*, 2017), while other known EPS are composed of maximum 4 different monosaccharides (Rehm, 2010). The degradation of WH15EPS would require a broader range of enzymes than other EPS, therefore we hypothesized that the application of WH15EPS to topsoil-litter samples would promote the enrichment of a wider range of GHs. We performed Stable Isotope Probing (SIP) combined with metagenomics, and coupled solid culture medium with metagenomics using WH15EPS as an enrichment factor to detect and cultivate potential GH producers and find GH genes with biotechnological potential.

2. Material and Methods

2.1. Soil samples

Four topsoil-litter mixed samples were collected in the spring of 2017 from the Wolfheze forest in the Netherlands (Table S1). Samples were taken from topsoil (0 to 5 cm) adjacent to fallen tree trunks. The collected samples were pooled, sieved (2 mm mesh) and immediately used for SIP incubation with EPS from *Granulicella* sp. strain WH15 (WH15EPS). The physicochemical properties of the topsoil-litter samples were determined (Eurofins Agro BV, Wageningen, NL) and are presented in Table S2.

2.2. SIP metagenome

The methods used for WH15EPS labeling, purification, sample collection, incubation and DNA extraction and fractionation are described in chapter 5. Library preparation and high-throughput shotgun sequencing were performed using the "heavy" DNA fractions pooled within each sample replicate as well as the total DNA of both the ^{12}C-EPS-amended and unamended controls. Shotgun sequencing was performed in 4 replicates of samples collected after 35 days of incubation (latest timepoint). Library preparation and Illumina MiSeq PE250 shotgun sequencing were performed at McGill University and Génome Québec Innovation Centre (Montréal, Québec, Canada). The sequences were deposited in the European Nucleotide Archive (ENA; https://www.ebi.ac.uk/ena) under the accession number PRJEB31257.

2.3. Metagenome of cultivated microorganisms grown in culture media with WH15EPS as sole carbon source

For evaluation of the metagenome of microorganisms that were able to grow in culture medium with WH15EPS as a sole carbon source, ten grams of fresh topsoil-litter sample were mixed with 100 ml of 100 mM MES buffer (2-[N-morpholino]ethanesulphonic acid, 1.95 g/l, pH 5.5), agitated for 30 min at room temperature on a vortex and decanted for 30 min. Dilutions (10^{-3} to 10^{-6}) were prepared in sterile MES buffer and 200 µl of the dilutions were plated in

quadruplicate. Diluted culture medium DNMS [$MgSO_4.7H_2O$ 0.2g/l, $CaCl_2.2H_2O$ 0.053g/l, chelated iron solution 0.2 ml/l (ferric III ammonium citrate 0.1g/100 ml, EDTA 0.2g/100ml, HCl 0.3ml/100ml) trace element solution SL10 1ml/L (Atlas, 2010), NH_4Cl 0.1g/l, agar 20g/l] with added WH15EPS (Kielak *et al.*, 2017) (0.05%) pH 5.5 and 40ng/µl (40 mg/l) cicloheximide to prevent growth of fungi was used for plating. To prevent caramelization, the freeze-dried purified WH15EPS was hydrated with milli-Q water, sterilized by filtration through a 0.2 µm membrane (Millipore) and added to the culture medium after autoclaving. Chelated iron solution and trace element solution SL10 were added after autoclaving and cooling of the culture medium. The plates inoculated with the soil suspension were incubated at room temperature for 1 month. The dilution 10^{-3} was chosen for sequencing. After incubation, colonies were scraped and used for total DNA extraction with PowerSoil® DNA Isolation Kit (MO BIO Laboratories, Inc). Following the first DNA extraction, a second round of DNA extraction was performed for each sample, according to Dimitrov et al. (2017). The total DNA extracted from the plates was used for metagenome shotgun sequencing. Library preparation and Illumina HiSeq XTen sequencing were performed at Genewiz (Suzhou, China). The sequences were deposited in the European Nucleotide Archive (ENA; https://www.ebi.ac.uk/ena) under the accession number PRJEB24069.

2.4. Bioinformatics and statistical analyses of metagenome data

2.4.1. SIP metagenome

SIP metagenome sequences were processed using EBI MGnify (Mitchell *et al.*, 2017) pipeline and SqueezeMeta (Tamames & Puente-Sánchez, 2019) pipeline in sequential mode. Briefly, in the SqueezeMeta pipeline, trimming and quality filtering were performed using Trimmomatic (Bolger *et al.*, 2014), assembly for each sample separately was done using Megahit (Li *et al.*, 2015); Prodigal (Hyatt *et al.*, 2010) was used for Open Reading Frame (ORF) prediction, and barrnap (Seemann, 2018) was employed for small subunit (SSU) rRNA gene sequence retrieval, which were classified using RDP classifier (Wang *et al.*, 2007). Diamond (Buchfink *et al.*, 2014) software was used for taxonomic classification of the ORFs against Genbank nr database and functional annotation with eggNOG database, for KO and COG numbers (Huerta-Cepas *et al.*, 2016). eggNOG-mapper (Huerta-Cepas *et al.*, 2017) was employed for carbohydrate-active enzymes annotation with against dbCAN database (Yin *et al.*, 2012). SqueezeMeta script SQM2tables.py was used to compute the average coverage and normalized TPM (transcripts per million) values for information on gene and function abundances. Normalized TPM SqueezeMeta ORF dataset and SSU rRNA gene data recovered from MGNify analysis were used for statistical analyses, performed in RStudio version 1.1.423 running R version 3.5.1 (R Core Team, 2015). For the SSU rRNA gene-based analysis, OTUs with less than 1 count across all the samples, chloroplast and mitochondrial sequences were discarded; prior to alpha diversity analyses, the data were rarefied to the size of the smallest sample (175 reads). For both ORF-based and SSU rRNA gene-based taxonomy datasets, 'Phyloseq' package

(McMurdie & Holmes, 2013) was used to calculate the number of observed OTUs, Shannon and Inverse Simpson diversity indices, and Chao1 and ACE diversity estimators. Bray-Curtis distance matrices constructed using the Hellinger transformed (Legendre & Gallagher, 2001) datasets were used for principal coordinate analysis (PCoA) using the capscale function from the 'vegan' package v. 2.4.6 (Oksanen *et al.*, 2018). Group dissimilarities were tested by permutational multivariate analysis of variance (PERMANOVA) using the function Adonis from the 'vegan' package. CANOCO (version5) (Braak & Smilauer, 2012) was employed to explore the relationship between sample treatments and taxa abundance through redundancy analysis (RDA) in the Hellinger transformed datasets. The statistical significance (p-value <0.05) of eigenvalues and treatment-taxa abundance correlations were tested using Monte Carlo permutation test at 499 permutations and the top 20 taxa associated with the dispersion of the treatments were displayed in RDA graphs.

In order to identify predicted functions (COG, KEGG and CAZymes) responsible for the observed clustering patterns, we performed a feature selection using a 'random forest' algorithm using the R package Boruta (Kursa & Rudnicki, 2010) (1,000 trees, p-value <0.05). Boruta tests if the importance of each individual variable is significantly higher that the importance of a random variable by fitting random forest models iteratively until all predictor variables are classified as "confirmed" or "rejected" at the 0.05 alpha level (Leutner *et al.*, 2012). The heatmaps for each function were constructed with pheamap (Kolde, 2019) R package, based on z-score transformed TPM (transcripts per million) abundances to improve normality and homogeneity of the variances. Sequences were submitted to the European Nucleotide Archive (ENA) and are available under the accession number PRJEB31257.

2.4.2. Metagenome of cultivated microorganisms

The DNA of the cultivated microorganisms were shotgun metagenome sequenced and the sequences were processed using EBI MGnify (Mitchell *et al.*, 2017) pipeline and ATLAS (Automatic Tool for Local Assembly Structures) (White III *et al.*, 2017) pipeline. For ATLAS, quality filtering was performed using BBDuk2 and cross-assembly was done with Megahit (Li *et al.*, 2015); functional and taxonomic analysis were performed at ORF level for the assembled contigs. Prodigal (Hyatt *et al.*, 2010) was used for ORF prediction and eggNOG database (Huerta-Cepas *et al.*, 2016) was used for functional annotation (COG and KO numbers) using DIAMOND software (Buchfink *et al.*, 2014). EggNOG-mapper (Huerta-Cepas *et al.*, 2017) was used for functional annotation of CAZymes with dbCAN database (Yin *et al.*, 2012). Kaiju software (Menzel *et al.*, 2016) was used for ORF taxonomy assignment against NCBI RefSeq database. Custom scripts were used to generate tables containing information of taxonomy and function abundance of the ORFs in all samples. Quality controlled contigs >1000 kb were used for binning using Concoct (Alneberg *et al.*, 2014), Maxbin (Wu *et al.*, 2014) and Metabat (Kang *et al.*, 2015); resulting bins were refined using DAS tool(Sieber *et al.*, 2018) and genome dereplication was performed with dRep (Olm *et al.*, 2017). Completeness and

contamination of the assembled genomes were checked using CheckM (Parks *et al.*, 2015), as well as taxonomy assignment. The ORFs of the genomes were predicted using Prodigal (Hyatt *et al.*, 2010), and DIAMOND software (Buchfink *et al.*, 2014) was used for functional annotation with eggNOG (COG and KO numbers) (Huerta-Cepas *et al.*, 2016). The annotation of CAZymes was performed with EggNOG-mapper (Huerta-Cepas *et al.*, 2017) against dbCAN database (Yin *et al.*, 2012). Sequences were submitted to the European Nucleotide Archive (ENA) and are available under the accession number PRJEB24069.

3. Results

3.1. SIP metagenome

3.1.1. Overview of the metagenome data

After quality control filtering, a total of 18,762,958 reads were maintained for further analysis, with an average of 1,563,580 reads per sample. A total of 1,209,745 ORFs were predicted for functional annotation, and approximately 50% of these ORFs were classified using KEGG and COG databases. The sequencing statistics are in Table 1.

Table1: SIP shotgun metagenomics sequencing statistics for each treatment. Average from 4 replicates.

Sequence statistics	Unamended Control	EPS amended	"Heavy" fraction
Number of reads	1,590,046	1,591,447	1,509,247
Number of contigs	82,370.25	78,474.5	92,521.75
Longest contig (bp)	2,651	3,645.25	9,414.75
N50	466	473	542
Mapping (%)	18.4	19.6	32.9
Number of ORFs	96,141	91,272	115,023.3

3.1.2. Community composition based on SSU rRNA and ORF classification

Taxonomic annotation based on SSU rRNA annotation demonstrated that bacteria, fungi and archaea accounted for approximately 84%, 4% and 2% of the sequences, respectively. At phylum level, 17 bacterial groups, 5 fungal groups and 3 archaeal groups were observed in all the samples. The most abundant groups at phylum level belonged to kingdom *Bacteria* (Figure S1a). *Proteobacteria* was the most abundant phylum in all treatments (26.4-28% of the sequences), followed by *Actinobacteria* (14.5-17.5% of the sequences). In both unamended and ^{12}C-EPS-amended control treatments, *Acidobacteria* was the third most abundant group (14.5-15.8% of the sequences), while in the "heavy" fraction samples, *Planctomycetes* was the third most abundant phylum (16.45% of the sequences) (Figure S1a). At genus level, we observed 167 groups in all samples, of which 110 were unclassified groups. "Unclassified *Bacteria*" was the most abundant group in the unamended control (3.5% of the sequences),

while "unclassified *Acidobacteriaceae*" (2.6% of the sequences) was the most abundant in the ^{12}C-EPS-amended control (Figure 1a). In labeled samples, the predominant group was "unclassified *Planctomycetes*" (3.2% of the sequences) (Figure 1a). Among the 10 most abundant groups, only 2 classified genera were observed: *Acidothermus* (1.8-2.9% of the sequences) and *Singulisphaera* (0.2-2.6% of the sequences) (Figure 1a). Similarly, the taxonomic composition of the ORF-based analysis was dominated by kingdom *Bacteria*, with an average of 82% of the ORFs belonging to bacteria and approximately 18% of the ORFs

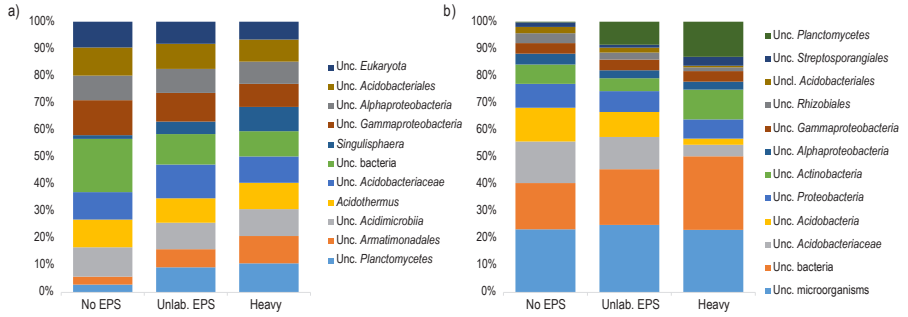

Figure 1: Taxonomic composition and relative abundance of microbial groups at genus level in SIP metagenome treatments based on a) SSU rRNA gene taxonomic classification b) ORF taxonomic classification. Only the ten most abundant groups for each treatment are displayed. Average abundances of 4 replicates. Unc: unclassified. No EPS – incubation without WH15EPS. Unlab. EPS-incubation containing ^{12}C-WH15EPS. Heavy – 'heavy fraction' of incubations containing ^{13}C-WH15EPS

originating from unclassified organisms, in all the samples (Figure S1b). At phylum level, we observed, in total, 103 bacterial groups, 6 fungal groups and 11 archaeal groups in all the samples. *Acidobacteria* (20.1-25.3% of the sequences) was the most abundant phylum in unamended and ^{12}C-EPS-amended control samples, while *Actinobacteria* (26% of the sequences) was the predominant group in "heavy" fraction samples (Figure S1b). At genus level, we found 1541 groups, of which 667 were unclassified. The top three most abundant groups in both control treatments were "unclassified microorganisms" (17.3-19.4% of the ORFs), "unclassified *Bacteria*" (12.7-16% of the ORFs) and "unclassified *Acidobacteriaceae*" (9.3-11.5%), while the predominant groups in "heavy" fraction samples were "unclassified microorganisms" (16.1% of the ORFs), "unclassified *Bacteria*" (18.9% of the ORFs) and "unclassified *Planctomycetes*" (9% of the ORFs) (Figure 1b).

PERMANOVA (p-values<0.001) showed that, for both SSU rRNA data and ORF based analysis, the microbial communities were different between treatments, with both control treatments closer to each other, and "heavy" fraction samples separated from both control treatments in PCoA graphs (Figure 2). For SSU rRNA communities, the first two axes of PCoA explained 43.3% of the variation, while for ORF based data, 90.6% of the variation was explained. RDA analysis for both datasets showed that mainly groups of *Planctomycetes,* such as "unclassified *Planctomycetes*", "unclassified *Planctomycetales"*, "unclassified *Planctomycetia"* and *Singulisphaera*, were driving the dispersion of the microbial communities between "heavy"

fraction and both control treatments (Figure S2), consistently with the higher abundance of *Planctomycetes* in labeled samples. Alpha diversity indices showed that richness and diversity indices were lower for "heavy" fraction samples in comparison with both controls (Figure S3), supported by ANOVA test (p-value < 0.05).

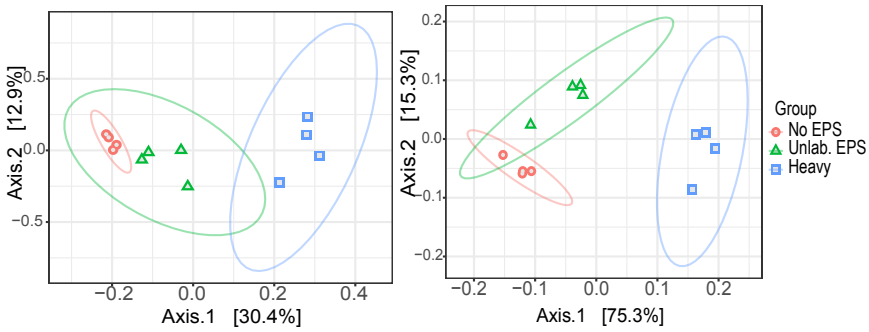

Figure 2:. Principal Coordinate Analysis (PCoA) clustering of normalized and Hellinger-transformed SIP metagenome sequencing data based on Bray-Curtis distances of a) SSU rRNA gene taxonomic classification and b) ORF taxonomic classification. No EPS – incubation without WH15EPS. Unlab. EPS-incubation containing ^{12}C-WH15EPS. Heavy – 'heavy fraction' of incubations containing ^{13}C-WH15EPS

3.1.3. Functional profile

KEGG, COG and dbCAN databases were employed for functional gene annotation to explore the functional characteristics of the microbial communities. Approximately 60% of the ORFs were assigned to COGs, matching in total to 20,644 COGs. The most abundant COG categories in all the samples were "R-General function prediction" (10.8-11.6% of the ORFs), "E-Aminoacid transport and metabolism" (10-5-11% of the ORFs), "C-Energy production and conversion" (8.1-8.6%), "L-Replication, recombination and repair" (7.1-7.6% of the ORFs) and "G-Carbohydrate transport and metabolism" (6.4-6.8%) (Figure S4a). Boruta feature selection "random forest" analysis was used to identify feature annotations that segregated significantly (p-value <0.05) between treatments. A total of 32 COGs were selected by Boruta algorithm. 13 among the identified COGs were more abundant in the unamended control samples, while 19 were more abundant in the labeled samples (Figure 3a). However, most of the features identified by the analysis belonged to the category unknown function. Some of the unknown COGs abundant in the labeled treatment, though, were associated mostly to phyla *Planctomycetes* and *Acidobacteria*, according to eggNOG database v 4.5 (Table S3).
KEGG analysis demonstrated that about 50% of the ORFs were assigned to 7,343 KEGG functional orthologs. The 17 most abundant KEGGs in all samples were assigned to three categories: signaling and cellular processes (8 KEGGs – 0.16% of the total ORFs), genetic information and processing (6 KEGGs – 0.14% of the total ORFs) and metabolism (3 – 0.21% of the total ORFs) (Figure S4b). Boruta feature selection identified 40 KEGGs that influenced the dispersion of the samples, of which 26 were more abundant in the labeled treatment and 14 were more abundant in the unamended control (Figure 3b). Among the KEGGs more

abundant in the labeled treatment, 13 could be assigned to KEGG pathways, mostly related to “metabolic pathways” and “microbial metabolism in diverse environments” (Table S4). Within the KEGGs more abundant in the unamended control treatment, 8 could be assigned to KEGG pathways, the majority related to “metabolic pathways” (Figure 3b, Table S4).

Annotation using dbCAN database showed that families GT41 (8.4%-11% of the CAZYmes), AA3 (4.4%-5%), GT4 (3.4%-4.7%), GT2 (4.1%-4.3%) and CE10 (3.5%-4.2%) were among the most predominant in all the treatments (Figure S4c). Boruta feature selection identified 27 CAZyme families affecting the dispersion of the sample treatments (Figure 3c), the vast majority belonging to the category glycoside hydrolase (GH). Among the selected families, 15 were more abundant in the labeled treatment and 12 were more abundant in the unamended control. The categories abundant in the labeled treatment involved xylan

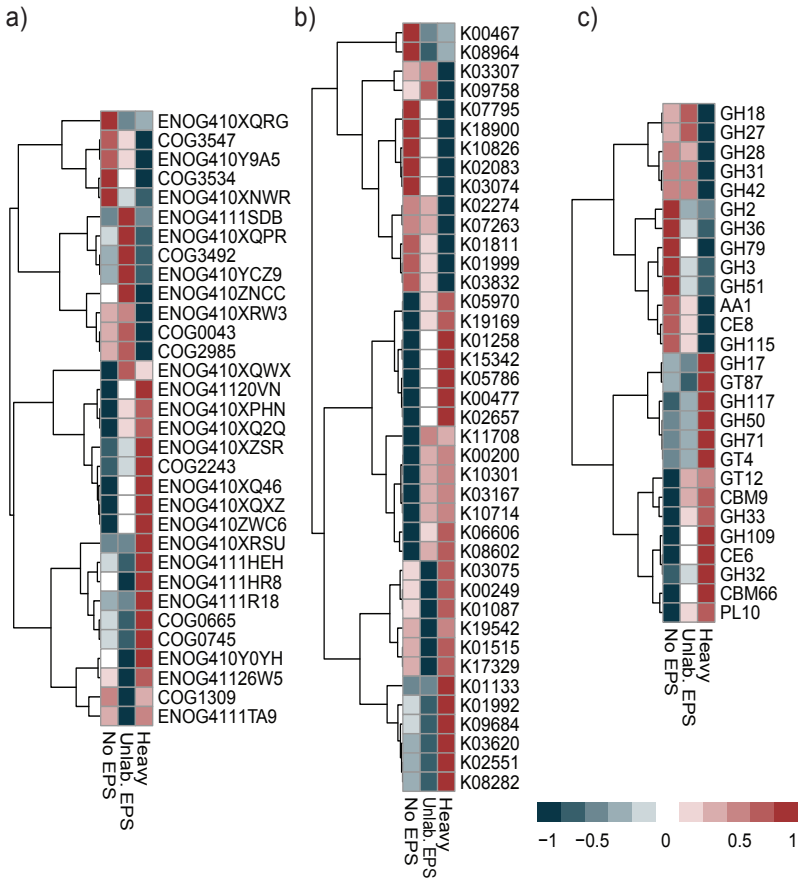

Figure 3: Boruta random forest feature selection of functions that significantly segregated across treatments based on 1000 permutations (p-value <0.05) for a) COG annotation, b) KEGG annotation and c) dbCAN annotation. Heatmaps based on the z-scored TPM normalized relative abundances of annotated ORFs from SIP metagenome samples. The description of the functions displayed in the heatmap are detailed in Table S3 (COG), Table S4 (KEGG) and Table S5 (dbCAN). No EPS – incubation without WH15EPS. Unlab. EPS-incubation containing ^{12}C-WH15EPS. Heavy – ‘heavy fraction’ of incubations containing ^{13}C-WH15EPS.

and fructan modules, xylanases, mannosytransferases and agarases, while the categories abundant in the unamended controls are mostly α and β galactosidases and glucosidases (Table S5). PERMANOVA (p-values<0.001) demonstrated that for KEGG, COG and dbCAN data, the functional gene compositions were different between treatments, similarly to taxonomic analysis, with control treatments grouping together and separated from "heavy" fraction samples (Figure S5).

3.2. Metagenome of cultivated microorganisms

3.2.1. Overview of the metagenomics data

A total of 422,735,048 reads were obtained after sequence quality 01 filtering, with an average of 80% of the ORFs classified with KEGG and COG databases. The sequencing statistics are described in Table 2.

Table 2: Sequencing statistics for shotgun metagenome of cultivated microorganisms. Average from 2 replicates per plate of culture medium.

Sequence statistics	Plate1	Plate2	Plate3	Plate4
Number of reads	49,148,370	54,247,258.5	58,397,852	49,574,043.5
Assembled reads	1.47^{E+10}	1.6224^{E+10}	1.75^{E+10}	1.4796^{E+10}
Number of contigs	67,980,868	76,125,070.5	82,202,170	68,767,888
Number of predicted genes	159,832	254,727	535,677	479,683
KEGG (% classified ORFs)	66.2	67.2	65.2	63.9
COG (%classified ORFs)	94.6	94.5	94.1	94.0
CAZYmes (%)	4.5	4.7	4.7	4.5
GC content (%)	59.6	60.3	59.0	58.2

3.2.2. Community composition based on SSU rRNA and ORF classification

Analysis of the taxonomic composition based on SSU rRNA showed an average of 73% of the sequences belonged to kingdom *Bacteria*, 20% to kingdom *Fungi* and 7% were derived from other Eukaryotes (Figure S6a). At phylum level, 17 bacterial groups, 7 fungal groups and 14 eukaryotic groups were identified. The most abundant group was the bacterial phylum *Proteobacteria*, with ~47.9% of the sequences, followed by fungal phylum *Ascomycota*, with ~ 14.5% of the sequences (Figure S6b). At genus level, 450 groups in total were observed, with the most abundant groups being bacterial groups. The predominant groups were "unclassified *Bacteria"* (~2.2% of the sequences) and *Dyella* (~1.5% of the sequences) (Figure 4a). *Silvimonas* and *Burkholderia* were also among the top 10 most abundant genera (~1.4 and 1.3% of the sequences, respectively). Similarly, for the ORF based data, the most abundant groups at genus level belonged to kingdom *Bacteria,* revealing the presence of 1930 groups at genus level. "Unclassified microbes" was the most abundant group, followed by genera *Caballeronia* (15.4% of the ORFs) and *Paraburkholderia* (15.1% of the ORFs) (Figure 4b). Other genera, such as *Burkholderia*, *Rhodanobacter* and *Dyella* were also among the predominant groups (7.8, 7.1 and 4.9% of the ORFs) (Figure 4b).

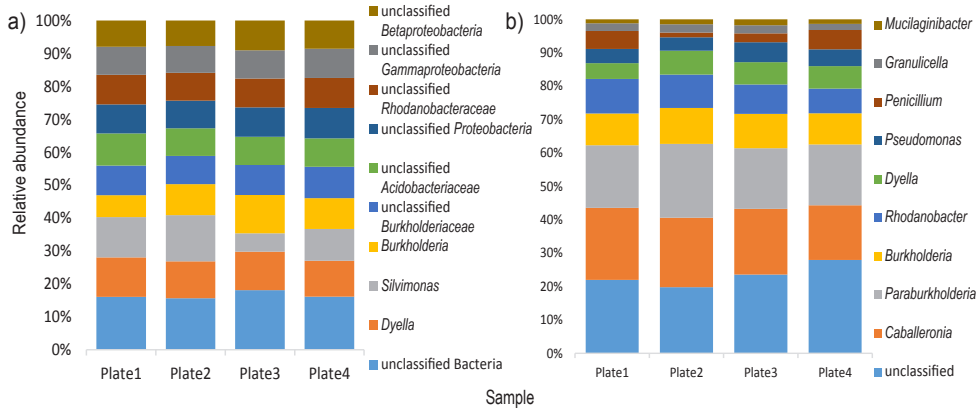

Figure 4: Taxonomic composition and relative abundance of microbial groups at genus level in samples from the metagenome shotgun of cultivated microrganims based on a) SSU rRNA gene taxonomic classification b) ORF taxonomic classification. Only the ten most abundant groups are displayed. Average from 2 replicates per plate of culture medium.

3.2.3. Functional profile

The functional profile of the cultivated microbes' metagenome was explored through the annotation with KEGG, COG and dbCAN databases. COG analysis demonstrated that approximately 20.6% of the annotated COGs were assigned to category "S-function unknown". Among the classified COGs, similarly to SIP metagenome, the predominant categories involved "E-Aminoacid transport and metabolism" (~8.6% of the ORFs), "G-Carbohydrate transport and metabolism" (~8.0% of the ORFs) and "C-Energy production and conversion" (~7.3% of the ORFs) (Figure 5a).

KEGG pathway analysis showed that around 65% of the ORFs were assigned to 9945 KEGG orthologs. The 20 most abundant KEGGs were distributed in the categories "Genetic information processing" (1 KEGG- ~0.24% of the total ORFs), "Metabolism" (4 KEGGs - ~1.18% of the total ORFs) and "Signaling and cellular processes" (15 –KEGGs -4.54% of the ORFs), of which 13 KEGGs were classified as transporters (Figure 5b).

The analysis of the carbohydrate-active enzymes with dbCAN database demonstrated the presence of 298 CAZyme families. Twenty-three families were predominant, which abundance was above 1%. Within the most abundant families, we observed 2 AA families (7.75% of the CAZymes), 1 CBM family, 4 CE families, 10 GH families and 6 GT families (Figure 5c). Those CAZyme families comprise mostly enzymes with cellulolytic (alpha-glucosidases, alpha fucosidases), hemicellulolytic (alpha-rhamnosidases, alpha-xylosidases, alpha-mannosidases, beta-galactosidases) and cell wall metabolism activities (N-acetylglucosaminyltransferases, alpha-N-acetylgalactosaminidases and peptidoglycan lyases) (Table S6). The most abundant family was GT 41 (Figure 5c), which encompasses UDP-GlcNAc: peptide β-N-acetylglucosaminyltransferases and UDP-Glc: peptide N-β-glucosyltransferases, enzymes envolved in protein glycosilation. Among the GH families, the

most abundant was GH13.

Among all 127 GH families found in both metagenome datasets, 114 families were observed in both datasets, while 5 families were exclusive from the SIP dataset (GH112, GH48, GH52, GH86, GH98) and 8 were exclusive from the cultivated microbial dataset (GH111, GH131, GH132, GH134, GH45, GH7, GH80, GH85) (Figure S7).

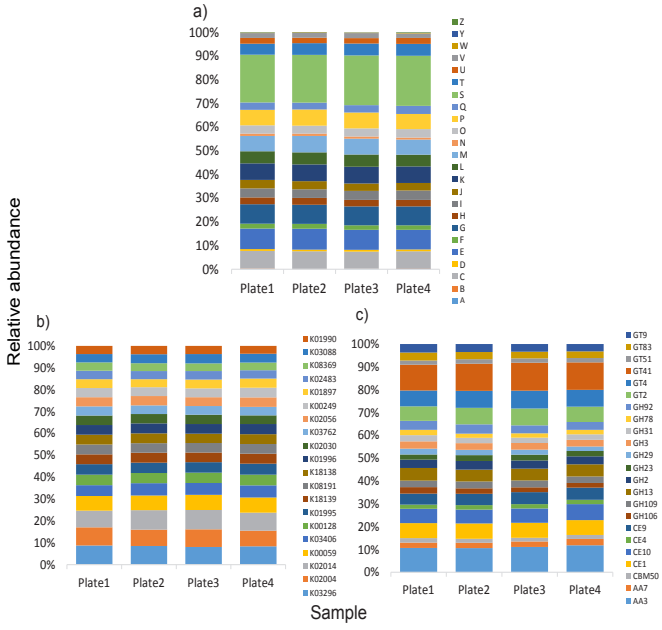

Figure 5: Relative abundance distribution of the most abundant functional categories in TMM-normalized metagenome sequencing data from the shotgun metagenome of cultivated microorganisms. a) COG annotation (all categories); b) KEGG annotation (above 0.2 % abundance); c) dbCAN annotation (above 1% abundance). The description of the functions displayed in b) and c) are detailed in Table S6. Average from 2 replicates per plate of culture medium. E-Amino acid transport and metabolism; G- Carbohydrate transport and metabolism; H-Coenzyme transport and metabolism; C-Energy production and conversion; I-Lipid transport and metabolism; F-Nucleotide transport and metabolism; Q- Secondary metabolites; D-Cell cycle; N-Cell motility; M-Cell wall/membrane/envelope biogenesis; V-Defence mechanisms; P-Inorganic ion transport and metabolism; U-Intracellular trafficking; O-Post translational modification; T-Signal transduction mechanisms; L-Replication, recombination and repair; K-Transcription; J-Translation; S-Function unknown; R-General function and prediction; X-Mobilome.

3.2.4. Taxonomy of the enriched glycoside hydrolase families

Taxonomic analysis of the most abundant GH family in both metagenome datasets, GH13, demonstrated that the majority of the sequences of GH13 in the cultivated microbes dataset belonged to phyla *Proteobacteria* (66.8% of the GH sequences) and *Acidobacteria* (21.8% of the GH sequences), while in the SIP dataset the most abundant phyla for GH13 were *Actinobacteria* (20.4-45.7% of the GH sequences), *Acidobacteria* (4-24.7% of the sequences) and other phyla (27-34% of the GH sequences) (Table3).

Table 3: Taxonomy associated to sequences of glycoside hydrolases belonging to GH13 family (most abundant) and the enriched GH families in heavy fraction samples from SIP metagenome.*

GH families	Sample	*Proteobacteria*	*Acidobacteria*	*Actinobacteria*	*Planctomycetes*	Others
GH13	Cultivated	66.8 (2605)	21.8 (827)	0.9 (36)	0.09 (1)	10.4 (51.4)
GH13_SIP	Control	20.1(128)	23.4 (148)	25.6 (162)	0.1(4)	31(196)
	EPS	17.9(89)	24.7(125)	20.4(95)	2.9(13)	34(172)
	Labeled	17.2(127)	4(30)	45.7(333)	5.7(42)	27(201)
GH109	Control	11(22)	45(92)	12(26)	2(4)	31(68)
	EPS	7(19)	26(76)	7(18)	21(60.8)	40(111.6)
	Labeled	7(34)	9(48)	13(62)	29(150)	42(217)
GH117	Control	0(0)	33(1)	33(1)	0(0)	33(1)
	EPS	0(0)	0(0)	17(1)	0(0)	38(5)
	Labeled	9(1)	0(0)	27(3)	0(0)	64(7)
GH50	Control	100(3)	0(0)	0(0)	0(0)	0(0)
	EPS	100(2)	0(0)	0(0)	0(0)	0(0)
	Labeled	8(2)	0(0)	0(0)	0(0)	92(24)
GH32	Control	0(0)	44(4)	11(1)	0(0)	44(4)
	EPS	0(0)	11(2)	5(1)	5(1)	79(15)
	Labeled	6(2)	14(5)	3(1)	9(3)	69(24)
GH17	Control	75(9)	0(0)	0(0)	0(0)	25(3)
	EPS	43(3)	0(0)	0(0)	0(0)	57(4)
	Labeled	44(8)	0(0)	0(0)	0(0)	56(10)
GH71	Control	0(0)	0(0)	100(1)	0(0)	0(0)
	EPS	0(0)	25(1)	50(2)	0(0)	25(1)
	Labeled	22(5)	0(0)	35(8)	0(0)	43(10)

*average percentage from 4 replicates (total number of sequences).

Within GH families that were more abundant in the SIP "heavy" fraction (Figure 4c), sequences of GH109 belonged mainly go *Acidobacteria* (45% of the GH sequences), other phyla (31-42% of the GH sequences) and *Planctomycetes* (2-29% of the GH sequences). GH117 family sequences belonged predominantly to *Actinobacteria* 17-33% of the sequences, *Acidobacteria* (0-33% of the GH sequences) and other phyla (33-64% of the GH sequences). Family GH50 sequences belonged mainly to *Proteobacteria* (8-100% of the GH sequences) and other phyla (0-92% of the GH sequences). GH 32 sequences were affiliated mainly to *Acidobacteria* (11-44% of the GH sequences) and other phyla (44-79% of the GH sequences). GH17 sequences belonged to phylum *Proteobacteria* (44-75% of the GH sequences) and other phyla (25-57% of the GH sequences). GH71 sequences were affiliated to phyla *Actinobacteria* (35-100% of the GH sequences), *Proteobacteria* (0-25% of the sequences), *Acidobacteria* (0-25% of the sequences) and other phyla (0-43% of the sequences).

3.2.5. Metagenome Assembled Genomes (MAGs) assembled from the metagenome of cultivated microorganisms

The binning process using contigs longer than 5 kb generated, after curation and quality filtering, 4 draft genomes. The genome length ranged from 3.0 to 6.3 Mb and the GC content ranged from 57 to 62%. All MAGs belonged to phylum *Proteobacteria*. None of the MAGs was classified to genus level, however the genomes were closer to genera *Paraburkholderia* (MAG1) and *Amantichitinum* (MAG2 and MAG4). MAG 3 closest classification was to family *Rhodanobacteraceae*. The characteristics of the genomes are described in Table 4. The

coverage of the genomes is described in Table S7.

Table 4: Genome characteristics for the 4 metagenome-assembled genomes (MAGs) obtained in this study.

Genome	MAG1	MAG2	MAG3	MAG4
Taxonomy (closest hit)	*Burkholderiaceae* 95% (*Paraburkholderia* 86%)	*Neisseriaceae* 42% (*Amantichitinum*: 42%)	*Rhodanobacteraceae* 77%	*Neisseriaceae*: 42% (*Amantichitinum* 42%)
Length (Mb)	6.3	3.0	4.8	3.7
Contigs	1997	997	80	1482
Completeness (%)	83.2	79.6	99.7	87.5
Contamination (%)	4.76	3.92	2.44	5
GC(%)	62	57	59	57
Number of predicted genes	7,126	3,580	4,280	4,552
Hits to protein database				
KEGG %	90.1	96.8	79.2	93.8
COG%	85.3	85.1	80.2	84
DBcan n(%)	279 (3.9)	141(3.9)	210 (4.9)	180 (4.0)

Approximately 83.7% of the ORFs predicted for the MAGs could be assigned to COGs. The analysis showed that most of the COG assigned ORFs fell on the category “S-function unknown” (16.4-18.4% of the ORFs). Among the classified COGs, however the most abundant categories were “K-transcription” (5.9-9% of the ORFs), “E-Amino acid metabolism” (4.8-8.1% of the ORFs), “G-Carbohydrate metabolism” (3.32-7.2%), “C-Energy production” (4.2-5.9%), “P-Inorganic ion metabolism” (4.65-6.3%) and “M-cell wall/membrane biogenesis” (5.2-5.9%) (Figure 6a).

KEGG pathway analysis demonstrated that around 90% of the predicted ORFs could be assigned to KEGG orthologs. The majority of the most abundant KEGG orthologs in all the MAGs were related to several types of transporter functions (Figure 6b and Table S8). In order to evaluate the features of the MAGs that could be involved in the uptake of the WH15EPS sugar units, we decided to look deeper into the transporters. Twenty-four of the KEGG orthologs observed in MAG1 genome were associated to the transport of several sugars, such as sorbitol, ribose, arabinose, xylose, fructose, rhamnose, glucose, mannose and multiple sugars (Table S9). Among the KEGG orthologs observed in MAG 2 genome, 62 were related to sugar transport, such as maltose, raffinose, lactose, glucosides, cellobiose, xylose, fructose, rhamnose, glucose, mannose and multiple sugars (Table S10). MAG3 did not exhibit sugar specific transporters within the 60 KEGGs related to transport function, however we observed some general type transporters (Table S11). In MAG4, 61 KEGG orthologs related to sugar transport were observed, such as maltose, raffinose, lactose, sorbitol, cellobiose, arabinose, xylose, fructose, rhamnose, glucose, mannose and multiple sugars (Table S12). We also performed the analysis of the CAZYmes with dbCAN database, in order to find enzymes that could be in associated the breakdown of the WH15EPS. MAG1 possessed 279 CAZymes

distributed in 90 families, of which the most abundant were CE1, GT4, GT42, CE10 and AA3 (Figure 6c). The seventy-six glycoside hydrolases observed were distributed in 43 families, including a wide range of activities, such as endo and exo-mannosidases, alpha and beta glucosidases and galactosidases, xylosidases, fucosidases and rhamnosidases (Table S13). MAG2 possessed 141 CAZymes distributed in 65 families and GT41, GT2 and CE1 were the most abundant families (Figure 6c). A total of 51 glycoside hydrolases from 30 families were observed, with activities such as alpha and beta glucosidases, beta galactosidases, mannanases and mannosidases, xylanases and polygalacturonases (Table S13). In MAG3, 210 cazymes distributed in 81 families were observed, and GT41, GT2, CE1 and CE10 were the most abundant (Figure 6c). Sixty-four glycosil-hidrolases distributed in 37 families were detected. The activities included alpha and beta galactosidases, alpha glucosidases, mannosidases, mannanases, rhamnosidases, arabinosidades, chitinases and trehalases

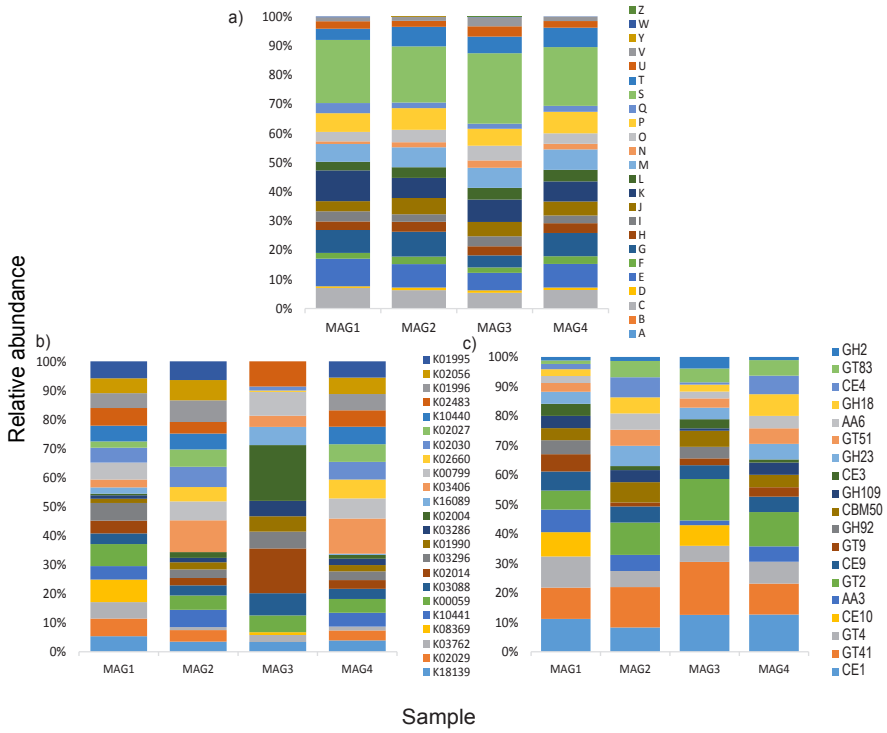

Figure 6: Relative abundance distribution of the most abundant functional categories in metagenome assembled genomes (MAGs) assembled from the shotgun metagenome of cultivated microorganisms sequencing data. a) COG annotation (all categories); b) KEGG annotation (top 10 for each genome); c) dbCAN annotation (top 10 for each genome). The description of the functions displayed in b) and c) are detailed in Table S8 and Table S13, respectively. E-Amino acid transport and metabolism; G- Carbohydrate transport and metabolism; H-Coenzyme transport and metabolism; C-Energy production and conversion; I-Lipid transport and metabolism; F-Nucleotide transport and metabolism; Q- Secondary metabolites; D-Cell cycle; N-Cell motility; M-Cell wall/membrane/envelope biogenesis; V-Defence mechanisms; P-Inorganic ion transport and metabolism; U-Intracellular trafficking; O-Post translational modification; T-Signal transduction mechanisms; L-Replication, recombination and repair; K-Transcription; J-Translation; S-Function unknown; R-General function and prediction; X-Mobilome.

(Table S13). The genome of MAG4 displayed 180 CAZymes distributed in 73 families, of which the most abundant were CE1, GT2 and GT41 (Figure 6c). The 64 glycoside hidrolases were spread among 34 families, including activities such as chitinases, arabinofuranosidases, alpha and beta glycosidases, mannosidases, cellulases, xylanases and polygaracturonases (Table S13). The distribution of most abundant CAZYmes and GH families in both metagenomics datasets and MAGs is depicted in Figure S8.

4. Discussion

In the present study, we applied culture-independent and culture-dependent techniques to evaluate microbial diversity and functions involved in the degradation of a microbial biopolymer, WH15EPS, focusing on enzymes of biotechnological interest. First, we compared the functional potential of the environment with and without the presence of WH15EPS, evaluating the taxonomic and functional enrichment induced by the addition of the biopolymer. To this end we used Stable Isotope Probing (SIP), which was followed by metagenomics to evaluate the functional potential of the microorganisms grown in culture medium with WH15EPS as the sole carbon source.

It was already observed that specific functions and activities can be selectively targeted through enrichment and specific techniques such as SIP. Enrichment with specific substrates allows the manipulation of the local microbial community prior to the metagenomic DNA extraction, increasing the prevalence of target functions (Ekkers *et al.*, 2011). For instance, Verastegui *et al.* (2014) incubated five ^{13}C labeled plant-derived carbon substrates (glucose, cellobiose, xylose, arabinose and cellulose) to diverse soils for the characterization of active soil bacterial communities and their glycosyl hydrolases. Furthermore, SIP has been used for bioremediation studies, involving the breakdown and metabolization of compounds such as methanol (Ginige *et al.*, 2004), phenol (Padmanabhan *et al.*, 2003) and biphenyl (Lee *et al.*, 2011).

SIP analysis demonstrated that in both SSU rRNA and ORF-based characterization, the phyla *Proteobacteria*, *Actinobacteria*, *Acidobacteria* and *Planctomycetes* were the most abundant in WH15EPS amended and unamended treatments. However, the addition of WH15EPS to the litter samples promoted an increase in the abundance of the phylum *Planctomycetes*, which was more evident in 'heavy' fraction samples, showing that *Planctomycetes* also play an active part in the degradation of WH15EPS. Furthermore, at genus level in the SSU rRNA based analysis, "unclassified *Planctomycetes*" and *Singulisphaera*, which belong to the same phylum, are the most abundant groups in the labeled treatment, while "unclassified *Planctomycetes*" is also among the most abundant in the ORF based analysis.

Proteobacteria, *Actinobacteria* and *Acidobacteria* are widely known to be involved in carbon-degradation processes. Several *Proteobacteria*, such as *Dyella*, *Mesorhizobium* and

Sphingomonas have been linked to glucose (Pinnell *et al.*, 2014) and cellulose assimilation (Haichar *et al.*, 2007), while the genus *Rhodanobacter* was implicated in the degradation of aromatic compounds (Song *et al.*, 2016). Members of *Actinobacteria* were already associated with cellulose, cellobiose and glucose hydrolysis (Schellenberger *et al.*, 2009, Bao *et al.*, 2019). *Acidobacteria* have mainly been linked to the degradation of hemicellulose, especially xylan, and genomic analyses showed the presence of genes associated to the degradation of a wide variety of polysaccharides (de Castro *et al.*, 2013, Kielak *et al.*, 2016). The glycolytic potential of the phylum *Planctomycetes* was demonstrated recently by Ivanova et al. (2017), using transcriptomics to evaluate their response to cellulose, xylan, pectin and chitin. The authors observed that each polymer induced an increase in abundance of transcripts belonging to different genera. Transcripts belonging to genus *Singulisphaera*, for instance, had their abundance increased significantly in response to pectin and xylan amendments.

The cultivation-dependent approach demonstrated, as expected, a lower taxonomic diversity, in which the widely studied *Proteobacteria* were among the most abundant. However, we also observed a wide diversity of unclassified bacteria and other unclassified microorganisms in both SSU and ORF-based analyses. The discrepancy between the diversity of taxa, especially the most abundant groups, observed in cultured and uncultured-based techniques is defined as “The Great Plate Count Anomaly” (Staley & Konopka, 1985). The cultivability of microorganisms in laboratory depends of many factors, such as nutrients, oxygen level, temperature, pH and growing factors (Vester *et al.*, 2015), limiting the total assortment of taxa that can be actually recovered in culture media. Nevertheless, adding WH15EPS as an alternative carbon source allowed us to demonstrate that several still unknown microorganisms can be grown in laboratorial conditions if unusual compounds are explored. The lower diversity in the culture media plates permitted the assembly of 4 draft genomes related to the most abundant *Proteobacteria*, which classification until genus level was not possible, once more demonstrating the enrichment and potential for isolation of previously unknown microbes.

In order to find potential enzymes of biotechnological interest we investigated the diversity of CAZymes in both culture-independent and culture-dependent generated datasets, due to their importance in almost all industrial sectors, such as chemical, pharmaceutical and food industries, as well as production of detergents, textiles, leather, paper and bioenergy (Berini *et al.*, 2017). Furthermore, we also investigated the presence of enzymes that could be employed for biofilm removal.

Among all CAZymes observed, the most abundant families belonged to Glycoside Transferase families, such as GT41, GT2 and GT4, either in culture based or in culture independent datasets. GTs are known to catalyze the formation of glycosidic bonds by transferring a sugar residue from a donor to an acceptor, which could be carbohydrates, proteins lipids, DNA and other molecules (Schmid *et al.*, 2016). Even though a large proportion of genes of microorganism`s genomes in general encode for GTs (abour 1-2% of the total number of genes) (Lairson *et al.*,

2008), those enzymes are still not as well explored as GHs (Schmid *et al.*, 2016). Glycosilated compounds play a wide range of roles, such as energy storage, cell integrity and signaling, among others, and the glycosilation of natural products is important in the exploration of bioactive compounds (Liang *et al.*, 2015). GTs are involved in the production of antibiotics, such as chloroeremomycin (Mulichak *et al.*, 2003), vancomycin (Mulichak *et al.*, 2004) and erythromycin D (Moncrieffe *et al.*, 2012), therefore they might be of interest especially for the pharmaceutical industry.

Within glycoside hydrolases, the most abundant family in both metagenomics datasets was GH13 (from *Proteobacteria*), which encompasses starch and pullulan modifying enzymes, including α-amylases, pullulanases, α-1,6-glucosidases, branching enzymes, maltogenic amylases, neopullulanases, and cyclodextrinases (Labes *et al.*, 2008). Amylases are among the most important enzymes for food industry, where they can be employed for production of glucose and maltose syrups, reduction of viscosity of syrups, production of clarified fruit juices, solubilization of starch for brewing processes and manufacture of baked products (Liu & Kokare, 2017). Furthermore, the application of α-amylases for the inhibition of biofilm formation has been investigated. In the study of Fleming *et al*. (2016) the use of amylase (from *Bacillus subtilis*) and cellulose (from *Aspergillus niger*) solutions to biofilms of *S. aureus* and *P. aeruginosa* decreased biomass significantly, increasing the effectiveness of antibiotics treatments. A similar effect was observed in the study of Craigen *et al*. (2011), where a commercially available α-amylase detached the aggregates produced by *S. aureus* and inhibited biofilm production.

Notwithstanding, feature selection with Boruta package revealed the differential abundance of GH families in "heavy" fraction SIP samples, originated from microorganisms that are believed to be able to degrade WH15EPS. These microorganisms belonged mainly to phyla *Proteobacteria*, *Acidobacteria*, *Actinobacteria*, *Planctomycetes*, as well as high proportion of unknown microorganisms. GH109 (*Acidobacteria* and *Planctomycetes*) contains α-N-acetylgalactosaminidases, which might be employed in the development of universal red blood cells, through the enzymatic removal of monosaccharides from red blood cells' membranes, and improvement of blood supply in hospitals (Liu *et al.*, 2007). Furthermore, those enzymes can be involved in the deconstruction of WH15EPS, since it contains units of xylose, glucose and arabinose (Kielak *et al.*, 2017). Families GH117 (*Acidobacteria* and *Actinobacteria*) and GH50 (*Proteobacteria*) contain agarases, which can be used for the production of oligosaccharides with antioxidant activities for applications in food, pharmaceutical and cosmetic industries (Fu & Kim, 2010). Family GH32 (*Acidobacteria*) comprises invertases and inulinases, enzymes that can be applied in food and fermentation processes (Khan *et al.*, 2013, Mohan *et al.*, 2018). GH17 (*Proteobacteria*) is composed of endoglucanases with activity against β-glucan and laminarin, effective additives for the degradation of polysaccharides for animal feed (Mohan *et al.*, 2018). Mutanases belonging to GH71 (*Actinobacteria*) family already showed activity against glucans present in dental

plaque (Wiater *et al.*, 2005).
Interestingly, sixteen of the most abundant GH families in the culture-independent dataset were found to be the predominant in the culture-dependent approach, and all the GH families with higher abundances in the labeled SIP samples were also observed in the culture-dependent dataset. Furthermore, the MAGs also contained GH families of interest. MAG 1 (similar to *Paraburkholderia*) contained 8 ORFs belonging to family GH92, which encompasses alpha mannosidases with applications in food and pharmaceultical industries, for the production of juices, degradation of plant material or coffee extraction (Konan *et al.*, 2016). In MAG2 (similar to *Amantichitinum*), five ORFs were classified as GH23, which contains lysozymes that can be used as polysaccharide hydrolysers for biofilm breakdown (Hukić *et al.*, 2018, Nahar *et al.*, 2018). MAG3 (*Rhodanobacteraceae*) is abundant in GH92 and GH23 but also GH2 family ORFs, which comprises several enzymes. Within the best characterized ones there are β-galactosidases employed for the production of lactose-free milk products and other galactooligosaccharides (Mallela *et al.*, 2016). MAG4 (similar to *Amantichitinum*) is rich in GH18 enzymes, envolving chitinases, that, for instance are important agents with applications for fungal biological control and bioremediation processes (Dahiya *et al.*, 2005). Our study showed that, using SIP and a complex EPS (WH15EPS), we could detect the subset of the total microbial community that was capable of incorporating the biopolymer. Among those we observed members of *Planctomycetes* as an interesting target for biotechnological studies and heterologous expression. In addition, we demonstrated that functional diversity induced by the presence of WH15EPS in both culture-dependent and culture independent approaches was enriched in genes coding for GHs, for instance, amylases, chitinases, agarases and endoglucanases and that could be applied in chemical, pharmaceutical and food industries. Furthermore, the use of WH15EPS may be employed for the investigation and isolation of yet unknown taxa, such as unclassified *Proteobacteria* and *Planctomycetes*, increasing the number of current cultured bacterial representatives.

Acknowledgements

We would like to thank Wietse de Boer for helping with the sampling local. O.Y.A. Costa was supported by an SWB grant from CNPq [202496/2015-5] (Conselho Nacional de Desenvolvimento Científico e Tecnológico).

Supplementary material

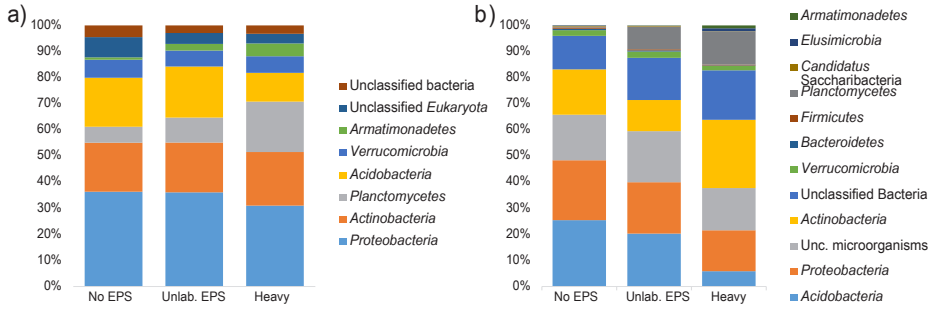

Figure S1: Taxonomic composition and relative abundance of microbial groups at phylym level in SIP metagenome treatments based on a) SSU rRNA gene sequence classification (>2.2 % abundance) b) ORF taxonomic classification (>0.1% abundance). Average abundances of 4 replicates. Unc: unclassified. No EPS – incubation without WH15EPS. Unlab EPS-incubation containing ^{12}C-WH15EPS. Heavy – 'heavy fraction' of incubations containing ^{13}C-WH15EPS.

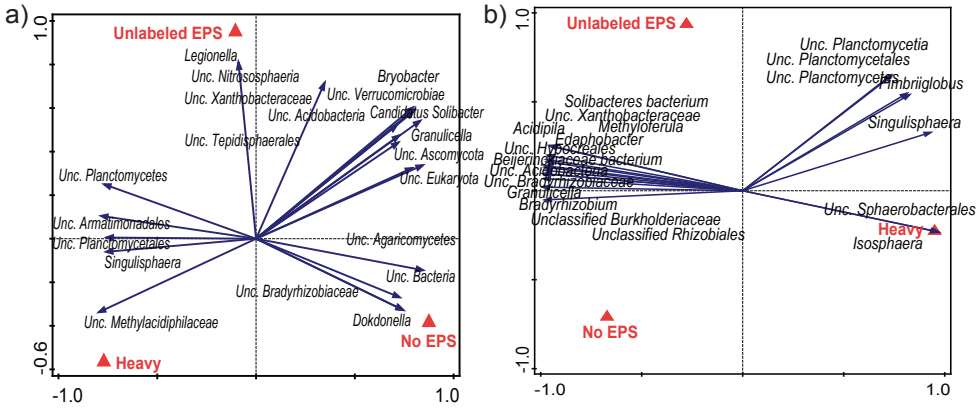

Figure S2: Biplot of the Redundancy analysis (RDA) based on normalized and Hellinger-transformed abundances of a) SSU rRNA gene taxonomy classification and b) ORF taxonomic classification. Only the best 20 fitting groups are displayed. Unc: unclassified. No EPS – incubation without WH15EPS. Unlab EPS-incubation containing ^{12}C-WH15EPS. Heavy – 'heavy fraction' of incubations containing ^{13}C-WH15EP.S

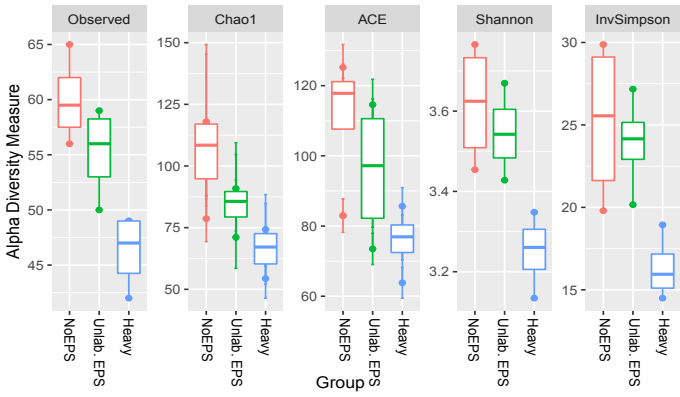

Figure S3: Box-plot comparisons of alpha-diversity assessment by richness estimators (number of observed OTUs, Chao1, ACE) and diversity indices (Shannon, Inverse Simpson) for SIP SSU rRNA gene samples. 'Heavy fraction' values are significantly lower in comparison with both controls for all comparisons (p-value < 0.05). Comparisons performed across treatments using ANOVA test and Tukey`s HSD post-hoc test. Data rarefied to the minimum sampling depth. Unlab. EPS-incubation containing ^{12}C-WH15EPS. Heavy – 'heavy fraction' of incubations containing ^{13}C-WH15EPS.

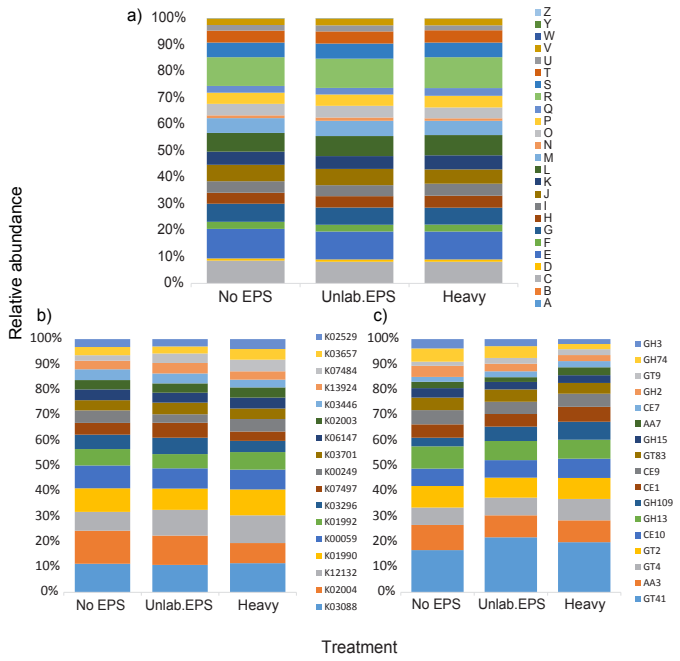

Figure S4: Relative abundance distribution of the most abundant functional categories in TPM-normalized metagenome sequencing data from the SIP metagenome. a) COG annotation (all categories); b) KEGG annotation (above 0.1 % abundance); c) dbCAN annotation (above 1% abundance). E-Amino acid transport and metabolism; G- Carbohydrate transport and metabolism; H-Coenzyme transport and metabolism; C-Energy production and conversion; I-Lipid transport and metabolism; F-Nucleotide transport and metabolism; Q- Secondary metabolites; D-Cell acycle; N-Cell motility; M-Cell wall/membrane/envelope biogenesis; V-Defence mechanisms; P-Inorganic ion transport and metabolism; U-Intracellular trafficking; O-Post translational modification; T-Signal transduction mechanisms; L-Replication, recombination and repair; K-Transcription; J-Translation; S-Function unknown; R-General function and prediction; X-Mobilome.

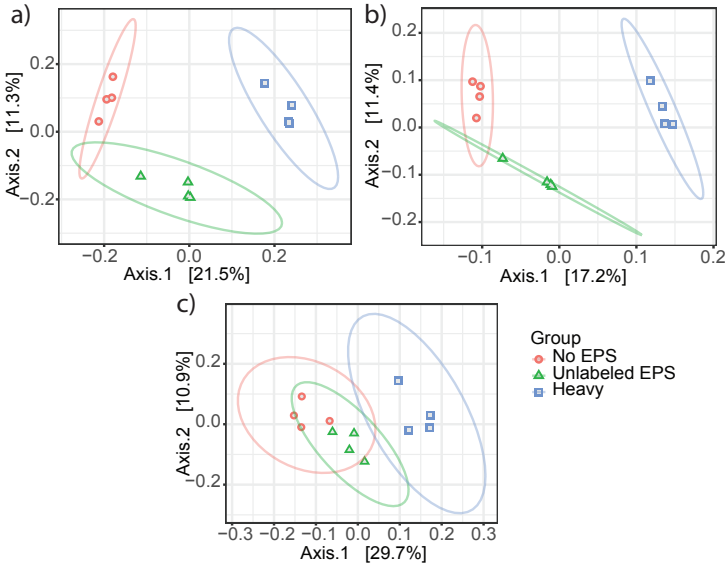

Figure S5: Principal Coordinate Analysis (PCoA) clustering of normalized and Hellinger-transformed SIP metagenome sequencing data based on Bray-Curtis distances of a) COG annotation, b) KEGG annotation and c) dbCAN annotation. No EPS – incubation without WH15EPS. Unlabeled EPS-incubation containing ^{12}C-WH15EPS. Heavy – 'heavy fraction' of incubations containing ^{13}C-WH15EPS.

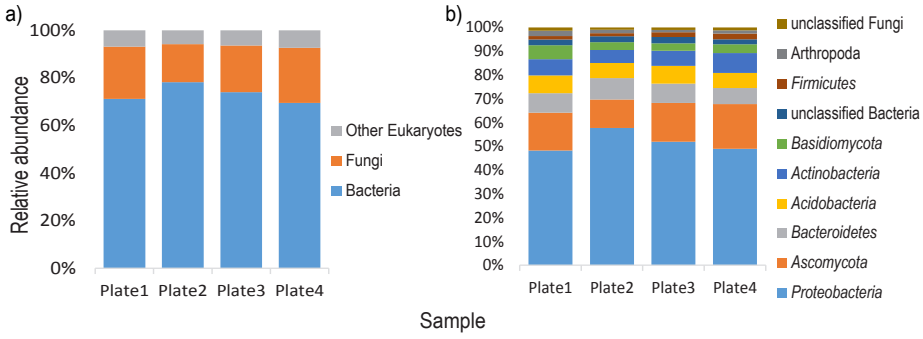

Figure S6: Taxonomic composition and relative abundance of microbial groups at a) kingdom and b) phylum level in samples from the metagenome shotgun of cultivated microrganims based on SSU rRNA gene taxonomic classification. Average from 2 replicates per plate of culture medium.

Figure S7: Venn diagram depicting the number of common and unique glycoside hydrolase (GH) families observed in SIP metagenome and metagenome of cultivate microorganisms` datasets.

a)

	Un.Control	EPS	Heavy
GT41	8.4%	11.1%	11.0%
AA3	5.0%	4.4%	4.8%
GT4	3.4%	3.6%	4.7%
GT2	4.3%	4.1%	4.6%
CE10	3.5%	3.6%	4.2%
GH13	4.4%	3.9%	4.2%
GH109	1.7%	2.9%	3.9%
CE1	2.6%	2.6%	3.3%
CE9	2.8%	2.4%	2.9%
GT83	2.5%	2.5%	2.4%
GH15	1.9%	1.5%	1.7%
AA7	1.3%	1.0%	1.7%
CE7	1.0%	1.2%	1.4%
GH2	2.2%	1.5%	1.3%
GT9	0.8%	1.2%	1.3%
CE14	1.0%	0.8%	1.3%
GH74	2.6%	2.4%	1.2%
GH3	1.9%	1.4%	1.1%
GT35	1.2%	0.9%	1.0%
GH33	0.1%	0.8%	1.0%

b)

	Plate1	Plate2	Plate3	Plate4
GT41	6.5%	6.9%	7.1%	6.9%
AA3	6.1%	6.1%	6.4%	6.8%
GT4	3.9%	4.3%	4.5%	4.2%
CE1	3.8%	3.9%	3.8%	3.7%
GT2	3.6%	4.2%	4.3%	3.8%
CE10	3.6%	3.5%	3.6%	4.0%
GH13	3.2%	3.0%	3.0%	3.0%
CE9	2.8%	2.9%	3.0%	3.0%
GH92	2.3%	2.4%	2.0%	1.9%
GT9	2.1%	2.0%	1.9%	1.8%
GH2	2.1%	2.2%	2.1%	2.0%
GT83	2.0%	1.8%	1.7%	1.7%
GH3	1.8%	1.7%	1.7%	1.7%
GH109	1.6%	1.7%	1.8%	1.6%
GH106	1.6%	1.4%	1.2%	1.2%
GH31	1.5%	1.4%	1.3%	1.3%
GH29	1.4%	1.4%	1.3%	1.1%
GH78	1.3%	1.1%	1.1%	1.1%
AA7	1.3%	1.3%	1.4%	1.6%
GH23	1.3%	1.5%	1.5%	1.3%

c)

	MAG1	MAG2	MAG3	MAG4
GT41	18	10	23	10
CE1	19	6	16	12
GT2	11	8	18	11
GT4	18	4	7	7
CE9	11	4	6	5
AA3	13	4	2	5
CBM50	7	5	7	4
CE10	14	0	9	0
GH23	7	5	5	5
GH18	4	4	3	7
GT51	5	4	4	5
GT83	2	4	6	5
GT9	10	1	3	3
CE4	3	5	1	6
AA3_1	6	3	2	4
AA6	4	4	3	4
GH109	7	3	1	4
GH3	5	3	4	3
CE3	7	1	4	1
GH92	8	0	5	0

d)

	Un.Control	EPS	Heavy
GH13	4.4%	3.9%	4.2%
GH109	1.7%	2.9%	3.9%
GH15	1.9%	1.5%	1.7%
GH2	2.2%	1.5%	1.3%
GH74	2.6%	2.4%	1.2%
GH3	1.9%	1.4%	1.1%
GH23	0.9%	1.1%	0.9%
GH78	0.9%	0.9%	0.8%
GH29	0.8%	1.0%	0.7%
GH92	0.8%	0.7%	0.7%
GH1	0.8%	0.7%	0.6%
GH28	1.1%	1.0%	0.5%
GH31	1.1%	1.0%	0.4%
GH39	0.9%	0.7%	0.4%
GH51	1.0%	0.7%	0.4%
GH106	1.1%	0.8%	0.4%
GH43	0.7%	0.6%	0.4%
GH127	0.8%	0.6%	0.4%
GH27	0.8%	0.9%	0.2%
GH144	0.7%	0.4%	0.2%

e)

	Plate1	Plate2	Plate3	Plate4
GH13	3.2%	3.0%	3.0%	3.0%
GH92	2.3%	2.4%	2.0%	1.9%
GH2	2.1%	2.2%	2.1%	2.0%
GH3	1.8%	1.7%	1.7%	1.7%
GH109	1.6%	1.7%	1.8%	1.6%
GH106	1.6%	1.4%	1.2%	1.2%
GH31	1.5%	1.4%	1.3%	1.3%
GH29	1.4%	1.4%	1.3%	1.1%
GH78	1.3%	1.1%	1.1%	1.1%
GH23	1.3%	1.5%	1.5%	1.3%
GH15	1.0%	0.9%	0.9%	0.9%
GH94	0.9%	1.2%	0.9%	0.9%
GH35	0.8%	0.8%	0.7%	0.7%
GH42	0.8%	0.7%	0.7%	0.6%
GH27	0.7%	0.9%	0.8%	0.7%
GH28	0.7%	0.8%	0.7%	0.7%
GH20	0.7%	0.8%	0.8%	0.7%
GH43	0.7%	0.6%	0.7%	0.8%
GH95	0.7%	0.7%	0.7%	0.6%
GH18	0.7%	0.7%	0.7%	0.7%

f)

	MAG1	MAG2	MAG3	MAG4
GH92	8	0	5	0
GH23	7	5	5	5
GH109	7	3	1	4
GH3	5	3	4	3
GH18	4	4	3	7
GH15	3	0	0	0
GH94	2	3	0	2
GH2	2	1	5	1
GH27	2	0	1	0
GH36	2	0	1	0
GH35	1	3	2	3
GH31	1	1	2	2
GH1	1	3	0	2
GH73	1	1	2	1
GH20	0	2	3	3
GH13_23	0	3	1	3
GH53	0	0	0	3
GH125	0	1	1	2
GH128	0	2	0	2
GH29	0	0	3	0

Figure S8: Distribution of the 20 most abundant CAZyme families in a) SIP metagenome samples (relative abundance, average of 4 replicates); b) Metagenome of cultivated microorganisms (relative abundance, average of 2 replicates); c) Metagenome-Assembled Genomes (MAGs) (number of genes), and most abundant glycosyl hydrolases (GH) in d) SIP metagenome samples (relative abundance, average of 4 replicates), e) metagenome of cultivated microorganisms (relative abundance, average of 2 replicates) and f) Metagenome-Assembled Genomes (MAGs) (number of genes).

Table S1: Coordinates of the sampling sites.

Site number	Latitude	Longitude
1	51°59'14.5"N	5°47'32.7"E
2	51°59'15.9"N	5°47'29.5"E
3	51°59'15.7"N	5°47'27.7"E
4	51°59'14.8"N	5°47'23.2"E

Table S2: Physicochemical properties of topsoil-litter samples.

Component	Unit	Average (Sd)	Component	Unit	Average (Sd)
total N	mg N/kg	16535±3217	**CEC**	%	81±0.0
C/N ratio		20±6	**CEC**	mmol+/kg	214±42
Available N	kg N/ha	252±187	**B**	µg B/kg	488±4.2
pH		3.05±0.1	**Cu**	µg Cu/kg	48±9.9
OM	%	55.8±3.5	**Fe**	µg Fe/kg	3070±466.7
Na	mg Na/kg	34.5±4.9	**Mn**	µg Mn/kg	101920±9362.1
P	mg P/kg	42.65±6.4	**Zn**	µg Zn/kg	8860±693.0
K	mg K/k	218±11.3	**Clay**	%	5.5±0.7
Ca	kg Ca/ha	13±0	**Silt**	%	10±14.1
Mg	mg Mg/kg	175±7.1	**Sand**	%	27.5±12.0

Sd: standard deviation

Table S3: COG functions that significantly segregated across treatments selected by Boruta random forests algorithm based on 1000 permutations in the SIP metagenome treatment comparisons (p-value <0.05).

COG	Function		Category
COG0043*	3-polyprenyl-4-hydroxybenzoate decarboxylase	H	Coenzyme transport and metabolism
COG2985*	transport protein/Uncharacterized membrane protein YbjL	P	Inorganic ion transport and metabolism
COG3492*	Deoxycytidine triphosphate deaminase	S	Function unknown
COG3534	Alpha-N-arabinofuranosidase (EC 3.2.1.55	G	G Carbohydrate transport and metabolism
COG3547*	Transposase	L	Replication, recombination and repair
ENOG410XNWR	ABC transporter substrate-binding protein	E	Amino acid transport and metabolism
ENOG410XQPR*	Sulfite exporter TauE/SafE	S	Function unknown
ENOG410XQRG	Transcriptional regulator	K	Transcription
ENOG410XRW3*	secreted protein	S	Function unknown
ENOG410Y9A5*	Cna B-type protein	S	Function unknown
ENOG410YCZ9*	NA	S	Function unknown
ENOG410ZNCC*	NA	S	Function unknown
ENOG4111SDB*	NA	S	Function unknown
COG0665	Glycine/D-amino acid oxidase	E	Amino acid transport and metabolism
COG0745	OmpR -regulatoR	T	Signal transduction mechanisms
COG1309	Transcriptional regulator	K	Transcription
COG2243	Precorrin-2 methylase	H	Coenzyme transport and metabolism
ENOG410XPHN*	NA	S	Function unknown
ENOG410XQ2Q*	sialic acid-specific 9-O-acetylesterase	S	Function unknown
ENOG410XQ46	Protein of unknown function (DUF1549) (*Planctomycetes*)	S	Function unknown
ENOG410XQWX*	NA	S	Function unknown
ENOG410XQXZ	NA	S	Function unknown
ENOG410XRSU	NA (*Acidobacteria*)	S	Function unknown
ENOG410XZSR	NA (*Acidobacteria*)	S	Function unknown
ENOG410Y0YH	Reductase	I	Lipid transport and metabolism
ENOG410ZWC6	Kelch repeat-containing protein	S	Function unknown
ENOG4111HEH	(ABC) transporter	S	Function unknown
ENOG4111HR8	Tetr family transcriptional regulator	K	Transcription

COG (continued)	Function		Category
ENOG4111R18	integral membrane transport protein	P	Inorganic ion transport and metabolism
ENOG4111TA9	NA	S	Function unknown
ENOG41120VN	NA (*Planctomycetes*)	S	Function unknown
ENOG41126W5	NA	S	Function unknown

*Higher abundance in ^{12}C-EPS-amended control samples. Blue – more abundant in unamended control samples. Orange: more abundant in Labeled samples

Table S4: KEGG orthologs that significantly segregated across treatments selected by Boruta random forests algorithm based on 1000 permutations in the SIP metagenome treatment comparisons (p-value <0.05).

KEGG	Annotation	KEGG Metabolic Pathway
K00467	Lactate 2-monooxygenase	Metabolic pathways/Pyruvate
K01811*	Alpha-D-xyloside xylohydrolase	-
K01999*	Branched-chain amino acid transport system substrate-binding protein	ABC transporters/quorum sensing
K02083	Allantoate deiminase	Metabolic pathways/Microbial metabolism/ Purines
K02274*	Cytochrome c oxidase subunit I	Metabolic pathways/Oxidative phosphorilation
K03074	Preprotein translocase subunit secf	Protein export/Bacterial secretion
K03307*	Solute:Na+ symporter, SSS family	-
K03832*	Periplasmic protein tonb	-
K07263*	Zinc protease	-
K07795	Putative tricarboxylic transport membrane protein	Two-component system
K08964	Methylthioribulose-1-phosphate dehydratase	Metabolic pathways/Amino-acids
K09758*	Aspartate 4-decarboxylase	Metabolic pathways/Amino-acids
K10826	Putative ABC transporter transmembrane protein	-
K18900	Lysr family transcriptional regulator, regulator for bpeef and oprc	-
K00200*	Formylmethanofuran dehydrogenase subunit A	Metabolic pathways/microbial metabolism/ carbon metabolism/methane metabolism
K00249	Acyl-coA dehydrogenase	Antibiotics/fatty acids/secondary metabolites
K00477	Phytanoyl-coA hydroxylase	Peroxysome
K01087	Trehalose 6-phosphate phosphatase	Metabolic pathways/starch sucrose met
K01133	Choline-sulfatase	-
K01258	Tripeptide aminopeptidase	-
K01515	ADP-ribose pyrophosphatase	Metabolic pathways/Purine metabolism
K01992	ABC-2 type transport system permease protein	-
K02551	2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylate synthase	Metabolic pathways/Ubiquinone biosynthesis
K02657	Twitching motility two-component system response regulator pilG	Two component system/Biofilm
K03075	Preprotein translocase subunit secG	Bacterial secretion/Protein export
K03167*	DNA topoisomerase VI subunit B	-
K03620	Ni/Fe-hydrogenase 1 B-type cytochrome subunit	Two componente system
K05786	Chloramphenicol-sensitive protein rarD	-
K05970*	Sialate O-acetylesterase	-
K06606*	2-keto-myo-inositol isomerase	Metabolic pathways/microbial metabolism/ inositol metabolism
K08282	Non-specific serine/threonine protein kinase	-
K08602*	Oligoendopeptidase F	-
K09684	Pucr family transcriptional regulator, purine catabolism regulatory protein	-
K10301*	F-box protein 21	-
K10714*	Methylene-tetrahydromethanopterin dehydrogenase	Metabolic pathways/ Microbial metabolism/ Carbon metabolism/ Methane metabolism

KEGG (continued)	Annotation	KEGG Metabolic Pathway
K11708*	Manganese/zinc/iron transport system permease protein	ABC transporters
K15342	CRISP-associated protein Cas1	-
K17329	N,N'-diacetylchitobiose transport system substrate-binding protein	ABC transporters
K19169*	DNA sulfur modification protein dndB	-
K19542	MFS transporter, DHA2 family, tetracycline/ oxytetracycline resistance protein	-

*Higher abundance in EPS amended samples. Blue – more abundant in unamended control samples. Orange: more abundant in Labeled samples

Table S5: CAZyme families that significantly segregated across treatments selected by Boruta random forests algorithm based on 1000 permutations in the SIP metagenome treatment comparisons (p-value <0.05).

CAZY family	Function
AA1*	Multicopper oxidases
CE8*	Pectin methylesterase
GH18*	Chitinase
GH2	β-galactosidase
GH27*	α-galactosidase
GH28*	Pectin polygalacturonase
GH3	β-glucosidase/xylosidase
GH31*	α-glucosidases
GH36	α-galactosidase/N-acetylgalactosaminidase
GH42*	β-galactosidases
GH51	Hemicellulases
GH79	Proteoglycanases
GH115*	Xylan α-glucuronidase
CBM66	Fructan-binding modules
CBM9*	Xylan-binding modules
CE6	Acetyl xylan esterase
GH109	α-N-acetylgalactosaminidase
GH117	α-1,3-L-neoagarooligosaccharide hydrolase
GH17	1,3;1,4-β-D-glucan endohydrolases
GH32	Invertases
GH33*	Sialidases
GH50	β-agarase
GH71	α-1,3-glucanase
GT12*	[N-acetylneuraminyl]-galactosylglucosylceramide N-acetylgalactosaminyltransferase
GT4	Mannosyltransferases/N-acetylglucosaminyltransferases
GT87	Polyprenol-P-Man: α-1,2-mannosyltransferase
PL10*	Pectate lyase

AA-auxiliary activity; CE-carbohydrate esterase; GH-glycoside hydrolase; GT-glycoside transferase; PL-polysaccharide lyases. *Higher abundance in EPS-amendedsamples. Blue – more abundant in unamended control samples. Orange: more abundant in Labeled samples

Table S6: Most abundant CAZyme families (above 1% abundance) and most abundant KEGG orthologs (above 0.2% abundance) in the shotgun metagenome of cultivated microorganisms.

Category	Ave abundance (%)	Associated functions
CAZy family		
GT41	6.83%	UDP-GlcNAc: peptide β-N-acetylglucosaminyltransferase; UDP-Glc: peptide N-β-glucosyltransferase
AA3	6.35%	Glucose-methanol-choline (GMC) family of oxidoreductases,
GT4	4.24%	Mannosyltransferases/ N-acetylglucosaminyltransferases
CE1	3.78%	Large variety of substrates

Category (continued)	Ave abundance (%)	Associated functions
GT2	3.96%	Cellulose synthase, chitin synthase, mannosyltransferase, glucosyltransferase, galactosyltransferase, rhamnosyltransferase
CE10	3.67%	Esterases acting on non-carbohydrate substrates
GH13	3.05%	α-glucosidase
CE9	2.93%	Deacetylation of N-acetylglucosamine-6-phosphate to glucosamine-6-phosphate
GH92	2.15%	α-mannosidase
GT9	1.97%	Lipopolysaccharide N-acetylglucosaminyltransferase/Heptosyltransferase
GH2	2.07%	β-galactosidase
GT83	1.80%	Undecaprenyl phospho-α-L-4-amino-4-deoxy-L-arabinose; dodecaprenyl phospho-β-galacturonic acid:lipopolysaccharide core α-galacturonosyl transferase
GH3	1.71%	β-glucosidases/galactosidases/xylosidases
GH109	1.69%	α-N-acetylgalactosaminidase
GH106	1.34%	α-L-rhamnosidase/rhamnogalacturonan α-L-rhamnohydrolase
GH31	1.40%	α-xylosidase/glucosidase/mannosidase
GH29	1.29%	α-fucosidase
GH78	1.16%	α-L-rhamnosidase/rhamnogalacturonan α-L-rhamnohydrolase
AA7	1.40%	Gluco/chito-oligosaccharide oxidase
GH23	1.39%	Peptidoglycan lyase
CBM50	1.02%	Bind to the N-acetylglucosamine residues in bacterial peptidoglycans and in chitin
GT51	1.08%	Murein polymerase
CE4	1.08%	De-acylation of polysaccharides
KEGG		
K03296	0,50%	Hydrophobic/amphiphilic exporter-1 (mainly G- bacteria), HAE1 family
K02014	0,50%	Iron complex outermembrane recepter protein
K02004	0,46%	Putative ABC transport system permease protein
K00059	0,41%	3-oxoacyl-[acyl-carrier protein] reductase
K03406	0,32%	Methyl-accepting chemotaxis protein
K01995	0,29%	Branched-chain amino acid transport system ATP-binding protein
K00128	0,28%	Aldehyde dehydrogenase (NAD+)
K18139	0,27%	Outer membrane protein, multidrug efflux system
K01996	0,27%	Branched-chain amino acid transport system ATP-binding protein
K18138	0,26%	Multidrug resistance, efflux pump MexAB-OprM
K08191	0,26%	MFS transporter, ACS family, hexuronate transporter
K02030	0,25%	Polar amino acid transport system substrate-binding protein
K02056	0,25%	Simple sugar transport system ATP-binding protein
K00249	0,25%	Acyl-CoA dehydrogenase
K03088	0,24%	RNA polymerase sigma-70 factor, ECF subfamily
K03762	0,24%	MFS transporter, MHS family, proline/betaine transporter
K01897	0,24%	Long-chain acyl-CoA synthetase
K02483	0,23%	Two-component system, OmpR family, response regulator
K01990	0,22%	ABC-2 type transport system ATP-binding protein
K08369	0,21%	MFS transporter, putative metabolite:H+ symporter
K03296	0,50%	Hydrophobic/amphiphilic exporter-1 (mainly G- bacteria), HAE1 family

Ave: average; AA-auxiliary activity; CE-carbohydrate esterase; GH-glycoside hydrolase; GT-glycoside transferase; PL-polysaccharide lyases

Table S7: MAGs coverage in all samples.

Sample	MAG1	MAG2	MAG3	MAG4
OCP1	160.7	18.9	1805.1	13.2
OCP2	126.8	29.4	1285.5	20.6
OCP3	252.2	81.7	1826.3	51.1
OCP4	119.9	58.3	1348.9	45.5
OCP5	191.6	0.0	1715.5	0.0
OCP6	127.2	48.0	1443.8	40.0
OCP7	150.3	0.0	1376.4	0.0
OCP8	178.4	90.3	1149.0	54.5

Table S8: Most abundant KEGG orthologs in MAGs and their associated functions. A selection of the top 10 most abundant KEGG orthologs in each genome is displayed.

KEGG	Associated function	Category
K18139	Outer membrane protein, multidrug efflux system	Transporters
K02029	Polar amino acid transport system permease protein	Transporters
K03762	MFS transporter, MHS family, proline/betaine transporter	Transporters
K08369	MFS transporter, putative metabolite:H+ symporter	Transporters
K10441	ribose transport system ATP-binding protein	Transporters
K00059	3-oxoacyl-[acyl-carrier protein] reductase	Lipid metabolism
K03088	RpoE	Transcription machinery
K02014	Iron complex outermembrane recepter protein	Transporters
K03296	Hydrophobic/amphiphilic exporter-1 (mainly G- bacteria), HAE1 family	Transporters
K01990	ABC-2 type transport system ATP-binding protein	Transporters
K03286	OmpA-OmpF porin, OOP family	Transporters
K02004	Putative ABC transport system permease protein	Transporters
K16089	Outer membrane receptor for ferrienterochelin and colicins	Transporters
K03406	Methyl-accepting chemotaxis protein	Signal transduction
K00799	Glutathione S-transferase	Metabolism of amino acids
K02660	Twitching motility protein PilJ	Secretion system
K02030	Polar amino acid transport system substrate-binding protein	Transporters
K02027	Multiple sugar transport system substrate-binding protein	Transporters
K10440	Tibose transport system permease protein	Transporters
K02483	Two-component system, OmpR family, response regulator	Two-component system
K01996	Branched-chain amino acid transport system ATP-binding protein	Transporters
K02056	Simple sugar transport system ATP-binding protein	Transporters
K01995	Branched-chain amino acid transport system ATP-binding protein	Transporters

Table S9: Sugar transporters in MAG1 annotated with eggNOG database.

KEGG	Gene	Type of transporter
K10111	*malK*	Multiple sugar transport
K10112	*msmX*	Multiple sugar transport
K10227	*smoE*	Sorbitol/mannitol transport system
K10228	*smoF*	Sorbitol/mannitol transport system
K10229	*smoG*	Sorbitol/mannitol transport system
K10439	*rbsB*	Ribose transport system
K10440	*rbsC*	Ribose transport system
K10441	*rbsA*	Ribose transport system
K10537	*araF*	L-arabinose transport system
K10538	*araH*	L-arabinose transport system
K10539	*araG*	L-arabinose transport system
K10543	*xylF*	D-xylose transport system
K10544	*xylH*	D-xylose transport system
K10545	*xylG*	D-xylose transport system
K10552	*frcB*	Fructose transport system
K10553	*frcC*	Fructose transport system
K10554	*frcA*	Fructose transport system
K10559	*rhaS*	Rhamnose transport system
K10560	*rhaP*	Rhamnose transport system
K10561	*rhaQ*	Rhamnose transport system
K10562	*rhaT*	Rhamnose transport system
K17315	*gtsA*	Glucose/mannose transport system
K17316	*gtsB*	Glucose/mannose transport system
K17317	*gtsC*	Glucose/mannose transport system

Table S10: Sugar transporters in MAG2 annotated with eggNOG database.

KEGG (continued)	Gene	Type of transporter
K06726	*rbsD*	D-ribose pyranase
K10108	*malE*	Maltose/maltodextrin transport system
K10109	*malF*	Maltose/maltodextrin transport system
K10110	*malG*	Maltose/maltodextrin transport system
K10111	*malK*	Multiple sugar transport
K10112	*msmX*	Multiple sugar transport
K10117	*msmE*	Raffinose/stachyose/melibiose transport system
K10118	*msmF*	Raffinose/stachyose/melibiose transport system
K10119	*msmG*	Raffinose/stachyose/melibiose transport system
K10191	*lacK*	Lactose/L-arabinose transport system
K10193	*togM*	Oligogalacturonide transport system
K10228	*smoF*	Sorbitol/mannitol transport system
K10229	*smoG*	Sorbitol/mannitol transport system
K10234	*aglG*	Alpha-glucoside transport system
K10235	*aglK*	Alpha-glucoside transport system
K10236	*thuE*	Trehalose/maltose transport system
K10237	*thuF*	Trehalose/maltose transport system
K10238	*thuG*	Trehalose/maltose transport system
K10240	*cebE*	Cellobiose transport system
K10241	*cebF*	Cellobiose transport system
K10242	*cebG*	Cellobiose transport system
K10439	*rbsB*	Ribose transport system
K10440	*rbsC*	Ribose transport system
K10441	*rbsA*	Ribose transport system
K10537	*araF*	L-arabinose transport system
K10538	*araH*	L-arabinose transport system
K10539	*araG*	L-arabinose transport system
K10540	*mglB*	Methyl-galactoside transport system
K10541	*mglC*	Methyl-galactoside transport system
K10542	*mglA*	Methyl-galactoside transport system
K10543	*xylF*	D-xylose transport system
K10544	*xylH*	D-xylose transport system
K10545	*xylG*	D-xylose transport system
K10548	-	Putative multiple sugar transport
K10550	*alsC*	D-allose transport system
K10552	*frcB*	Fructose transport system
K10553	*frcC*	Fructose transport system
K10554	*frcA*	Fructose transport system
K10558	*lsrA*	AI-2 transport system
K10559	*rhaS*	Rhamnose transport system
K10560	*rhaP*	Rhamnose transport system
K10561	*rhaQ*	Rhamnose transport system
K10562	*rhaT*	Rhamnose transport system
K15770	*cycB*	Arabinogalactan oligomer /
K15771	*ganP*	Arabinogalactan oligomer /
K15772	*ganQ*	Arabinogalactan oligomer /
K17204	*eryE*	Erythritol transport system
K17207	*xltA*	Putative xylitol transport
K17208	*ibpA*	Inositol transport system
K17209	*iatP*	Inositol transport system
K17210	*iatA*	Inositol transport system
K17213	-	Inositol transport system substrate-binding
K17214	-	Inositol transport system permease
K17215	-	Inositol transport system ATP-binding
K17241	*aguE*	Alpha-1 4-digalacturonate transport
K17242	*aguF*	Alpha-1 4-digalacturonate transport
K17244	*chiE*	Putative chitobiose transport
K17245	*chiF*	Putative chitobiose transport
K17246	*chiG*	Putative chitobiose transport

KEGG (continued)	Gene	Type of transporter
K17315	*gtsA*	Glucose/mannose transport system
K17316	*gtsB*	Glucose/mannose transport system
K17317	*gtsC*	Glucose/mannose transport system

Table S11: General type transporters in MAG3 annotated with eggNOG database.

KEGG	Type of transporter
K06147	ATP-binding cassette subfamily B bacterial
K02004	Putative ABC transport system permease
K02003	Putative ABC transport system ATP-binding
K01992	ABC-2 type transport system permease
K01990	ABC-2 type transport system ATP-binding
K03286	OmpA-OmpF porin OOP family
K11085	ATP-binding cassette subfamily B bacterial
K16013	ATP-binding cassette subfamily C bacterial

Table S12: Sugar transporters in MAG4 annotated with eggNOG database.

KEGG	Gene	Type of transporter
K06726	*rbsD*	D-ribose pyranase
K10108	*malE*	Maltose/maltodextrin transport system substrate-binding protein
K10109	*malF*	Maltose/maltodextrin transport system permease protein
K10110	*malG*	Maltose/maltodextrin transport system permease protein
K10111	*malK*	Multiple sugar transport system ATP-binding
K10112	*msmX*	Multiple sugar transport system ATP-binding
K10117	*msmE*	Raffinose/stachyose/melibiose transport system substrate-binding protein
K10118	*msmF*	Raffinose/stachyose/melibiose transport system permease protein
K10119	*msmG*	Raffinose/stachyose/melibiose transport system permease protein
K10191	*lacK*	Lactose/L-arabinose transport system ATP-binding protein
K10193	*togM*	Oligogalacturonide transport system permease protein
K10227	*smoE*	Sorbitol/mannitol transport system substrate-binding protein
K10228	*smoF*	Sorbitol/mannitol transport system permease protein
K10229	*smoG*	Sorbitol/mannitol transport system permease protein
K10234	*aglG*	Alpha-glucoside transport system permease protein
K10235	*aglK*	Alpha-glucoside transport system ATP-binding protein
K10236	*thuE*	Trehalose/maltose transport system substrate-binding protein
K10237	*thuF*	Trehalose/maltose transport system permease protein
K10238	*thuG*	Trehalose/maltose transport system permease protein
K10240	*cebE*	Cellobiose transport system substrate-binding protein
K10241	*cebF*	Cellobiose transport system permease protein
K10242	*cebG*	Cellobiose transport system permease protein
K10439	*rbsB*	Ribose transport system substrate-binding protein
K10440	*rbsC*	Ribose transport system permease protein
K10441	*rbsA*	Ribose transport system ATP-binding protein
K10537	*araF*	L-arabinose transport system substrate-binding protein
K10538	*araH*	L-arabinose transport system permease protein
K10539	*araG*	L-arabinose transport system ATP-binding protein
K10540	*mglB*	Methyl-galactoside transport system substrate-binding protein
K10541	*mglC*	Methyl-galactoside transport system permease protein
K10542	*mglA*	Methyl-galactoside transport system ATP-binding protein
K10543	*xylF*	D-xylose transport system substrate-binding protein
K10544	*xylH*	D-xylose transport system permease protein
K10545	*xylG*	D-xylose transport system ATP-binding protein
K10548	-	Putative multiple sugar transport system
K10550	*alsC*	D-allose transport system permease protein
K10552	*frcB*	Fructose transport system substrate-binding protein
K10553	*frcC*	Fructose transport system permease protein

KEGG (continued)	Gene	Type of transporter
K10554	*frcA*	Fructose transport system ATP-binding protein
K10559	*rhaS*	Rhamnose transport system substrate-binding protein
K10560	*rhaP*	Rhamnose transport system permease protein
K10561	*rhaQ*	Rhamnose transport system permease protein
K10562	*rhaT*	Rhamnose transport system ATP-binding protein
K15770	*cycB*	Arabinogalactan oligomer / maltooligosaccharide transport
K15771	*ganP*	Arabinogalactan oligomer / maltooligosaccharide transport
K15772	*ganQ*	Arabinogalactan oligomer / maltooligosaccharide transport
K17208	*ibpA*	Inositol transport system substrate-binding protein
K17213	-	Inositol transport system substrate-binding protein
K17214	-	Inositol transport system permease protein
K17215	-	Inositol transport system ATP-binding protein
K17241	*aguE*	Alpha-1 4-digalacturonate transport system substrate-binding
K17242	*aguF*	Alpha-1 4-digalacturonate transport system permease
K17244	*chiE*	Putative chitobiose transport system substrate-binding
K17245	*chiF*	Putative chitobiose transport system permease
K17246	*chiG*	Putative chitobiose transport system permease
K17313	*treU*	Trehalose transport system permease protein
K17315	*gtsA*	Glucose/mannose transport system substrate-binding protein
K17316	*gtsB*	Glucose/mannose transport system permease protein
K17317	*gtsC*	Glucose/mannose transport system permease protein
K17324	*glpS*	Glycerol transport system ATP-binding protein
K17325	*glpT*	Glycerol transport system ATP-binding protein

Table S13: Families of CAZymes observed in the MAGs, number of ORFs in each genome and associated enzymatic functions.

Family	MAG1	MAG2	MAG3	MAG4	Enzymatic activity
GH1	1	3	0	2	β-glucosidases and β-galactosidases:
GH10	0	1	0	1	Endo-beta-1,3-xylanase, endo-beta-1,4-xylanases
GH102	1	1	0	1	Lytic transglycosidease
GH103	1	1	1	1	Lytic transglycosideases
GH104	0	0	1	0	Lytic transglyosylases
GH105	0	0	1	0	Rhamnogalacturonidases
GH109	7	3	1	4	α-N-acetylgalactosaminidase
GH12	1	0	0	0	Endo-β-1,4-glucanase, endo-β-1,3-1,4-glucanase
GH123	0	0	1	0	N-acetyl-β-galactosaminidases
GH125	0	1	1	2	α-mannosidases
GH127	0	0	0	1	Arabinofuranosidase
GH128	0	2	0	2	β-1,3-glucanases
GH13_10	1	0	0	0	α-hydrolases, transglycosidases and isomerases
GH13_11	1	1	1	1	α-hydrolases, transglycosidases and isomerases
GH13_15	0	1	0	1	α-hydrolases, transglycosidases and isomerases
GH13_16	1	0	0	0	α-hydrolases, transglycosidases and isomerases
GH13_23	0	3	1	3	α-hydrolases, transglycosidases and isomerases
GH13_26	1	0	0	0	α-hydrolases, transglycosidases and isomerases
GH13_3	1	0	0	0	α-hydrolases, transglycosidases and isomerases
GH13_36	0	1	1	1	α-hydrolases, transglycosidases and isomerases
GH13_9	1	1	0	1	α-hydrolases, transglycosidases and isomerases
GH130	0	0	1	0	β-mannoside phosporilases
GH133	1	0	0	0	Amylo-α-1,6-glucosidase
GH135	1	0	0	0	Galactosaminogalactan hydrolase
GH141	1	0	0	0	α-L-fucosidase/xylanase
GH144	0	1	0	1	Endo-β-1,2-glucanase
GH15	3	0	0	0	Glucoamylase
GH16	1	0	0	0	β-1,4 or β-1,3 glycosidic bonds in various glucans and galactans 1,3-β-D-glucan endohydrolases and 1,3;1,4-β-D-glucan
GH17	0	0	2	0	Endohydrolases
GH18	4	4	3	7	Chitinases (EC 3.2.1.14) and endo-β-N-acetylglucosaminidase

Family (continued)	MAG1	MAG2	MAG3	MAG4	Enzymatic activity
GH19	0	0	1	0	Chitinases
GH2	2	1	5	1	β-galactosidases, β-glucuronidases, β-mannosidases, exo-β-glucosaminidases
GH20	0	2	3	3	Exo-acting β-N-acetylglucosaminidases, β-N-acetylgalactosamindase and β-6-SO3-N-Acetylglucosaminidases
GH23	7	5	5	5	Lytic transglycosideases
GH24	0	0	2	0	Lysozyme
GH27	2	0	1	0	α-Galactosidase
GH28	1	1	0	2	Polygalacturonases.
GH29	0	0	3	0	α-fucosidases
GH3	5	3	4	3	Exo-acting β-D-glucosidases, α-L-arabinofuranosidases, β-D-xylopyranosidases
GH30	0	0	1	0	β-glucosylceramidase, β-1,6-glucanase, and β-xylosidase
GH31	1	1	2	2	α-glucosidases
GH33	0	1	1	1	Sialidases and trans-sialidases
GH35	1	3	2	3	β-galactosidases
GH36	2	0	1	0	α-N-acetylgalactosaminidase
GH37	1	0	1	0	Trehalase
GH42	1	0	1	1	β-galactosidases, α-L-arabinosidase, β-D-fucosidase
GH49	1	0	0	0	dextranase, isopullulanase
GH5	1	0	0	0	Endo- and exoglucanases, endo- and exomannanases
GH5_13	1	0	0	0	Endo- and exoglucanases, endo- and exomannanases
GH5_19	0	1	0	1	Endo- and exoglucanases, endo- and exomannanases
GH5_43	1	0	0	0	Endo- and exoglucanases, endo- and exomannanases
GH5_48	0	1	0	1	Endo- and exoglucanases, endo- and exomannanases
GH50	0	0	1	0	β-agarase
GH51	0	0	0	1	L-arabinfuranosidases
GH53	0	0	0	3	β-1,4-galactanase
GH54	2	0	0	0	α-L-arabinofuranosidase and β-xylosidase
GH55	1	1	1	1	β-1,3-glucanases, including both exo- and endo
GH63	1	1	1	1	Exo-acting α-glucosidases
GH64	1	1	0	2	β-1,3-glucanase
GH65	0	0	2	0	Maltose phosphorylase, trehalose phosphorylase, kojibiose phosphorylase
GH73	1	1	2	1	β-N-acetylglucosaminidases.
GH76	0	1	1	0	Endo-acting α-mannanases
GH77	1	0	0	0	4-α-glucanotransferase
GH78	1	0	1	0	α-L-rhamnosidases
GH79	1	0	1	0	β-glucuronidase , β-4-O-methyl-glucuronidase , baicalin
GH87	1	0	0	0	β-glucuronidase , heparanase and hyaluronidase
GH9	0	0	1	1	α-1,3-glucanase
GH92	8	0	5	0	Cellulases
GH94	2	3	0	2	Exo-acting α-mannosidases, phosphorylase, cellodextrin phosphorylase, chitobiose phosphorylase
GH95	1	0	0	0	1,2-α-L-fucosidases, 1,2-α-L-galactosidases