



Article

The Structure of Stable Cellulolytic Consortia Isolated from Natural Lignocellulosic Substrates

Grigory V. Gladkov ^{1,*}, Anastasiia K. Kimeklis ^{1,2}, Alexey M. Afonin ¹, Tatiana O. Lisina ¹, Olga V. Orlova ¹, Tatiana S. Aksenova ¹, Arina A. Kichko ¹, Alexander G. Pinaev ¹ and Evgeny E. Andronov ^{1,3}

¹ All-Russian Research Institute of Agricultural Microbiology, 196608 Saint Petersburg, Russia

² Department of Applied Ecology, Saint-Petersburg State University, 199034 Saint Petersburg, Russia

³ Dokuchaev Soil Science Institute, 119017 Moscow, Russia

* Correspondence: ruginodis@gmail.com; Tel.: +7-921-402-65-16

Abstract: Recycling plant matter is one of the challenges facing humanity today and depends on efficient lignocellulose degradation. Although many bacterial strains from natural substrates demonstrate cellulolytic activities, the CAZymes (Carbohydrate-Active enZymes) responsible for these activities are very diverse and usually distributed among different bacteria in one habitat. Thus, using microbial consortia can be a solution to rapid and effective decomposition of plant biomass. Four cellulolytic consortia were isolated from enrichment cultures from composting natural lignocellulosic substrates—oat straw, pine sawdust, and birch leaf litter. Enrichment cultures facilitated growth of similar, but not identical cellulose-decomposing bacteria from different substrates. Major components in all consortia were from Proteobacteria, Actinobacteriota and Bacteroidota, but some were specific for different substrates—Verrucomicrobiota and Myxococcota from straw, Planctomycetota from sawdust and Firmicutes from leaf litter. While most members of the consortia were involved in the lignocellulose degradation, some demonstrated additional metabolic activities. Consortia did not differ in the composition of CAZymes genes, but rather in axillary functions, such as ABC-transporters and two-component systems, usually taxon-specific and associated with CAZymes. Our findings show that enrichment cultures can provide reproducible cellulolytic consortia from various lignocellulosic substrates, the stability of which is ensured by tight microbial relations between its components.

Keywords: lignocellulose decomposition; microbial consortium; CAZymes; glycoside hydrolases; metagenome sequencing; amplicon sequencing



Citation: Gladkov, G.V.; Kimeklis, A.K.; Afonin, A.M.; Lisina, T.O.; Orlova, O.V.; Aksenova, T.S.; Kichko, A.A.; Pinaev, A.G.; Andronov, E.E. The Structure of Stable Cellulolytic Consortia Isolated from Natural Lignocellulosic Substrates. *Int. J. Mol. Sci.* **2022**, *23*, 10779. <https://doi.org/10.3390/ijms231810779>

Academic Editor: Alessio Mengoni

Received: 17 August 2022

Accepted: 4 September 2022

Published: 15 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Plant biomass is one of the most abundant sources of organic carbon on the planet Earth, specifically due to the functioning of the modern industrial sectors [1]. Huge amounts of woody and herbaceous residues are generated as a byproduct of agriculture and forestry [2,3]. Most often these residues are considered waste; however, they can be reintegrated into the carbon cycle as a basis for the production of valuable products, such as biofuels, fertilizers, enzymes, organic acids, etc. [4].

The technical challenge of plant biomass handling lies in its chemical composition, the main component of which is lignocellulose [5], which, in turn, consists of three main compounds—cellulose, hemicellulose, and lignin, all highly recalcitrant to decomposition [6,7]. Other, less abundant compounds include proteins, pectin, soluble sugars, minerals, and lipids [8]. Their ratio differs between different lignocellulose-containing plant materials: cellulose makes up 9–80%, hemicellulose 10–50%, and lignin 5–35% [9]. The core of lignocellulose are cellulose microfibrils, surrounded by covalently linked molecules of hemicellulose and lignin. Cellulose is an unbranched homopolymer of glucose, while hemicellulose is a branched heteropolymer of different sugar residues, and lignin is an organic polymer derived from phenylpropane precursors [6,10].

Decomposition of lignocellulose is achieved by microorganisms using various multienzyme complexes. Cellulose processing is performed by three classes of enzymes: (1) endo- β -glucanases, which make cleavages in the internal bonds of the cellulose chain; (2) exo- β -glucanases, which cleave two (producing cellobioses) to four glucose residues from the cut in the chain; and (3) β -glucosidases (cellobioses), which convert cellobiose to glucose [11]. Hemicellulose decomposition requires additional specific enzymes, aimed at branching, acetylated sites, and different saccharide moieties, e.g., xylanases and mannanases [12]. These enzymes belong to the group of glycoside hydrolases (GHs) which catalyze the hydrolysis of glycosidic bonds [13]. GHs work in conjunction with Carbohydrate-binding modules (CBMs), which are covalently attached to the catalytic domains of the enzymes and promote interaction of the enzyme with the substrate [14]. Variability of GHs and CBMs ensures affinity to different sources of lignocellulose [15,16]. Other enzymes, which take part in a cellulolytic complex, are glycoside transferases (GT), which catalyze the transfer of saccharide moieties [17]. They work with a variety of donors and both saccharide and non-saccharide acceptors; thus, their diversity is also very high [18]. Lignin is degraded by a spectre of oxidoreductases, including laccases, high redox potential ligninolytic peroxidases, and oxidases [19]. All these enzyme categories are cataloged as modules in the Carbohydrate-Active enZymes (CAZy) Database, where they are classified into families [20]. These categories are highly diverse, e.g., GHs module consists of more than 100 families, defined more by sequence than by function [21]. Thus, enzymes specific for cellulose and hemicellulose decomposition are spread across many families of GHs. Enzymes for lignin decomposition are also present in CAZy's Auxiliary Activities (AAs) module [22].

Both bacteria and fungi participate in the process of lignocellulose decomposition, using GHs of the same families [23,24]. Fungal enzyme systems are well studied, but unlike bacteria, they are harder to cultivate, and they do not tolerate extreme environmental conditions; thus, they are costlier to use industrially [25]. Bacteria, on the other hand, are widely distributed in aerobic, anaerobic, saline, and extreme temperature habitats, and therefore can serve as a source of universal biodegraders. The most active members of cellulolytic community are found in Actinobacteria, Proteobacteria, Firmicutes, Chloroflexi, and Bacteroidetes phyla [26].

There are works focused on isolating cellulolytic bacterial strains from various biomass-containing environments, such as soils or plant and manure compost [27–29]. All these studies propose isolated strains as candidates for industrial lignocellulose conversion. However, in natural ecosystems' degradation of lignocellulose, biomass is carried out by multiple microorganisms with highly diverse spectres of enzymes, adaptive to different environmental conditions [30,31]. Thus, using bacterial consortia for biomass decomposition can overcome such difficulties in handling one-strain systems, such as negative feedback regulation and metabolite repression [32,33]. A way to isolate microbial consortium is by making enrichment cultures from natural communities using a specific carbon source [26,34,35]. This way the complex community of a natural ecosystem is funneled into simplified (with less variety) functional consortium specialized in this source [36]. Cortes-Tolalpa showed that microbial consortia derived from three different natural forest sources and enriched by wheat straw develop similar lignocellulolytic functional patterns while retaining unique taxonomic imprints [30]. However, not a lot of studies are focused on selecting reproducing minimal stable microbial consortia, which could be effectively saved and stored without altering their features.

In our study we aimed to isolate and explore stable cellulolytic consortia from enrichment cultures from composting plant biomass, common for northern regions of Europe—oat (*Avena*) straw, pine (*Pinus*) sawdust, and birch (*Betula*) leaf litter. These substrates represent different types of lignocellulose biomass—waste biomass (byproducts of agriculture and forestry) and virgin biomass (parts of naturally occurring plants), so our experiment aimed at incorporating different sources of common plant residues and testing their lignocellulolytic potential. We assessed the taxonomic structure and functional profiles of the

consortia derived from these substrates using methods of high-throughput sequencing of 16S rRNA gene amplicons and full metagenome analysis.

2. Results

As a result of the experiment on enrichment cultures from three different composting plant biomass sources, we acquired four reproducible microbial consortia with different phenotypic features, such as coloration of the filter paper and gas emissions (Figure S1). All consortia visually macerated filter paper (Figure 1). The specifications of the consortia are presented in Table 1. They covered all cellulose-containing substrates used in the experiment and represented some of the composting periods: consortia OS2 and OS4 from oat straw (2 and 4 months), consortium PS4 from pine sawdust (4 months), and consortium BL6 from birch leaf litter (6 months). Other combinations of substrates and composting periods did not result in reproducible consortia. Different composting substrates required different dilutions for the acquisition of minimal cellulolytic microbial cultures.

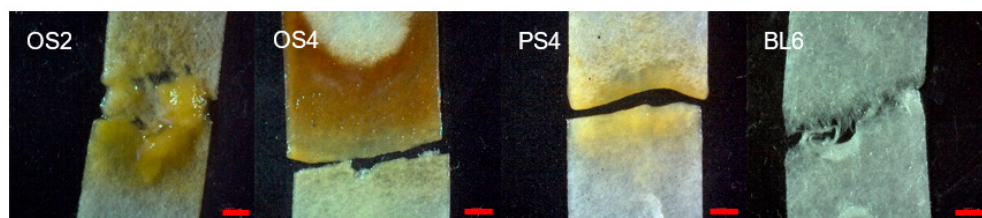


Figure 1. Maceration of the filter paper by each microbial consortium. OS2—oat straw, 2-month composting period; OS4—oat straw, 4 month; PS4—pine sawdust, 4 month; BL6—birch leaf litter, 6 month. The red scale bar is 2 mm.

Table 1. Characteristics of isolated microbial consortia.

Consortia ID	Substrate	Composting Time, Months	Substrate Dilution	Color of Consortia	Gas Formation
OS2	oat straw	2	10^6	yellow	no
OS4	oat straw	4	10^6	brown	yes
PS4	pine sawdust	4	10^3	yellow	no
BL6	birch leaf litter	6	10^4	transparent	yes

We aimed to address the microbial composition of isolated consortia by sequencing 16S rRNA gene and ITS amplicon libraries. Eukaryotes in all consortia were identified (both from amplicon and metagenome sequencing data) only as a minor component. The relative abundance of eukaryotes in the consortia according to metagenomic data was only 0.2%. For consortia OS2 and BL6, the presence of ciliates was shown (Figure S2). Consortia OS4 and BL6 had some proportion of unidentified fungi. Based on the obtained data, we can state that the isolated consortia are predominantly bacterial.

2.1. Taxonomic Analysis of Cellulolytic Consortia by Amplicon Sequencing

2.1.1. Alpha and Beta-Diversity of Consortia

As a result of amplicon sequencing, we acquired 24 libraries (six replicates for each of the four consortia) of 16S rRNA gene amplicons totaling 903,148 reads, which were divided into 535 phylotypes. All consortia demonstrated significant differences across the three indices of alpha-diversity (Figure 2). The highest values of all alpha-diversity indices were detected for the OS2 consortium, which was isolated from the substrate with the shortest composting period. Consortia OS4 and PS4, which were isolated from the substrates with 4-month composting periods, did not show significant differences in Simpson and Inverted Simpson indexes. Consortia BL6 had higher values of these indices than OS4 and PS4, but lower than OS2.

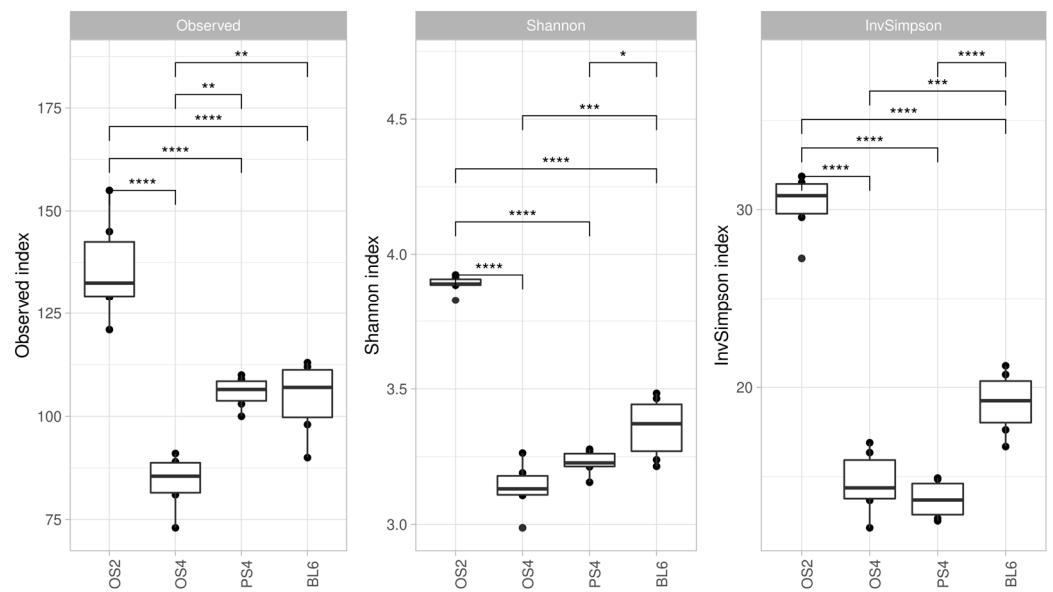


Figure 2. Alpha-diversity of four consortia (OS2, OS4, SD4, BL6) accessed by 3 indices—Observed, Shannon, InvSimpson, based on 16S rRNA gene libraries. Significant differences assessed by the Mann–Whitney test: (*) p -value ≤ 0.05 ; (**) p -value ≤ 0.01 ; (***) p -value ≤ 0.001 ; (****) p -value ≤ 0.0001 .

According to the beta-diversity metrics of 16s rRNA sequencing data, all four consortia were significantly separated from each other ($R^2 = 0.7$, p -value < 0.001) (Figure 3), even OS2 and OS4 from the same substrate ($R^2 = 0.85$, p -value = 0.0018). However, distances between consortia from oat straw were shorter than between others from different substrates. Bray–Curtis, UniFrac, and Weighted UniFrac metric systems showed the same trend but revealed differences in varying degrees (Figure S3).

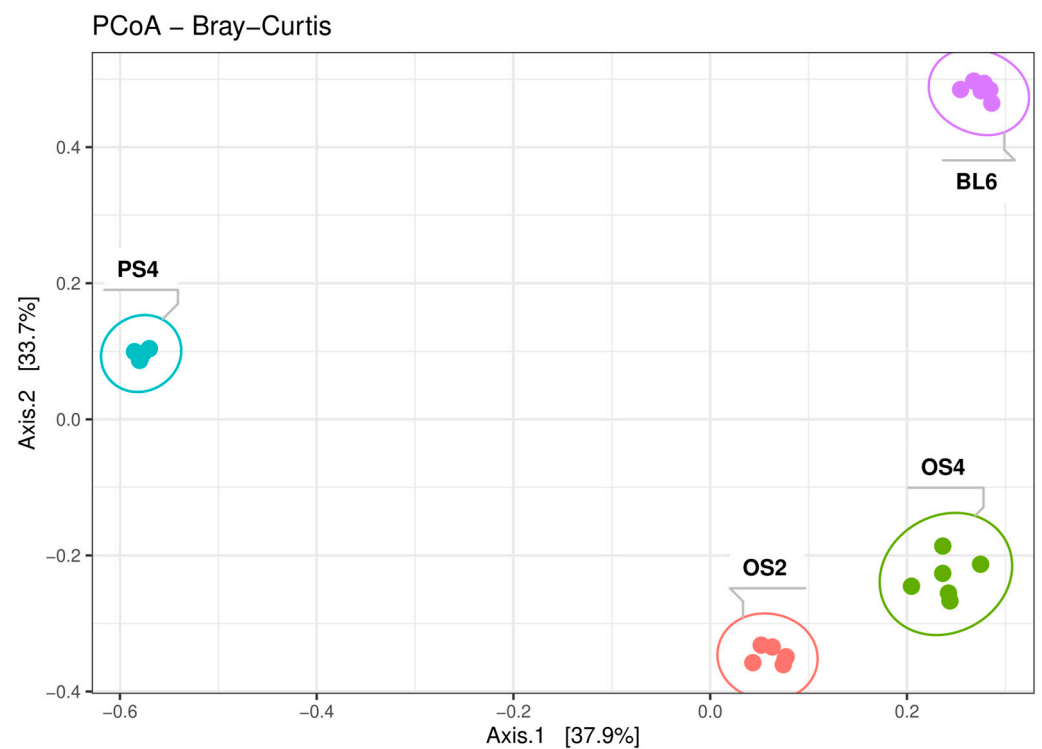


Figure 3. PCoA plot of the beta-diversity of the four cellulolytic consortia (OS2, OS4, PS4, BL6) accessed with Bray–Curtis, based on 16S rRNA gene amplicon libraries.

2.1.2. Taxonomic Composition of Consortia

According to amplicon sequencing, the main components of all consortia belonged to Proteobacteria and Bacteroidota phyla (Figure S4). To select the minimal cellulolytic community, consortia cultivated from the maximum dilution of the substrate, which retained the ability to decompose the paper filter, were taken into the study. Despite this, it turned out that each consortium consists of around a hundred phylotypes. Almost all of them were unique to different substrates, and only three minor phylotypes belonging to *Xanthobacter*, *Pseudoxanthomonas* and *Dokdonella* were common for all four consortia (Figure S5). However, phylotypes pooled on the genus level revealed overlapping representatives. Major genera can be seen on the heatmap with relative abundance data (Figure 4). *Pseudoxanthomonas* and *Devosia* genera from Proteobacteria were a major part in all consortia. Other phylotypes were distributed more differentially: *Sporocytophaga* (Bacteroidota) in OS2 and PS4; *Asticcacaulis* (Proteobacteria) in OS2, OS4, and PS4; *Cellulomonas* (Actinobacteroidota) in OS2, OS4, and BL6; *Caenimonas* and *Cellvibrionaceae* (Proteobacteria) in OS4; *Flavobacterium* (Bacteroidota) and *Blrii41* (Myxococcota) in OS2 and OS4; *Paenibacillus* (Firmicutes) and *Methyloversatilis* (Proteobacteria) in BL6; *Cohnella* (Firmicutes) in PS4 and BL6; *Pirellula* (Planctomycetota) in PS4.

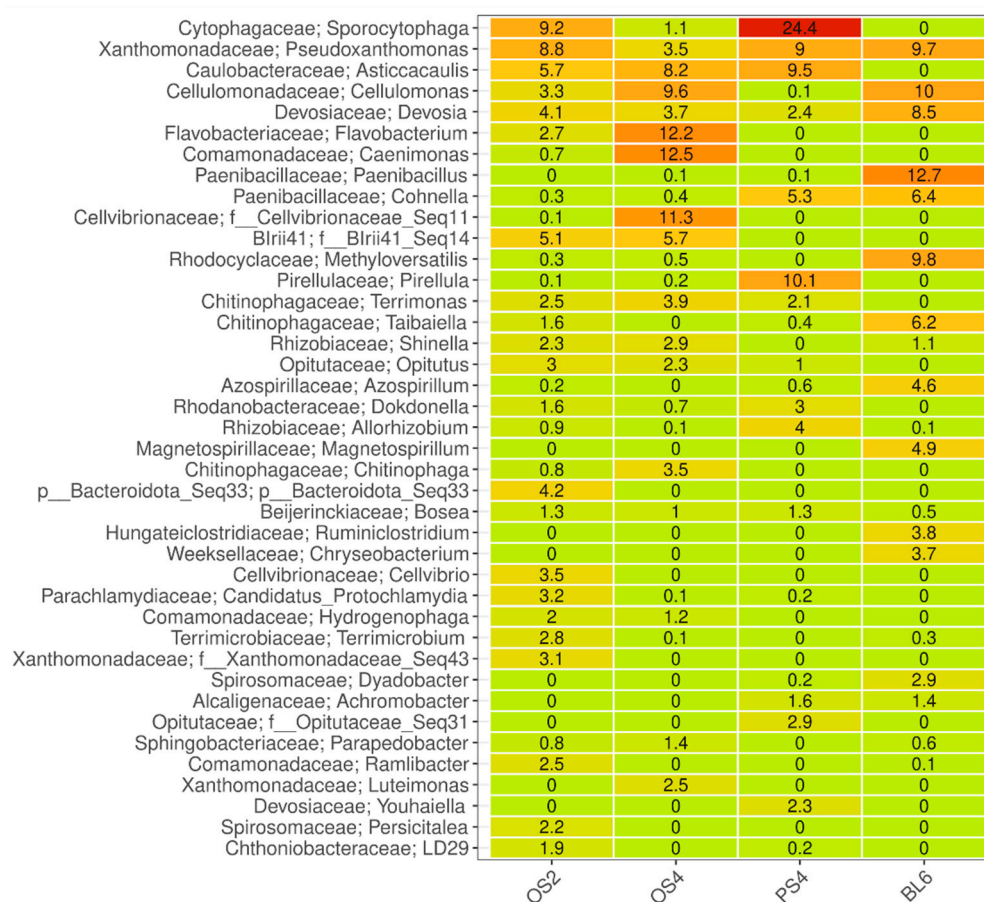


Figure 4. Heatmap of the most abundant bacterial phylotypes on the genus level in the four cellulolytic consortia (OS2, OS4, PS4, BL6). Relative abundance given in % of the read count of each consortium, red for maximal, and green for minimal values.

To overcome the limitations of relative abundance analysis, we compared phylotype composition between consortia by compositional PhILR-transformation analysis, which reveals associated shifts in taxon composition between each pair.

Consortia OS2 and OS4 from an oat straw substrate shared the most phylotypes, despite having different phenotypes: the number of the common phylotypes was 65, 68.4%

of the total read count from these consortia (Figure S6). Only 4% of the total reads were unique for OS4 in comparison with OS2, which had 26.8% of unique reads. The most abundant phylotypes from both consortia were similar, and consisted of representatives of Bacteroidetes, Gammaproteobacteria, Alphaproteobacteria, Myxococcota, Actinobacteria, and Verrucomicrobia. The main differences were at the low taxonomic level in members of Bacteroidota (Sphingobacteriaceae, between representatives of the genus *Parapedobacter*) and Proteobacteriota phyla (between phylotype groups within Comamonadaceae, Devosiaceae, Xanthomonadaceae, and *Bosea*). Combined data demonstrate that OS4 consortium differs from OS2 in reduced phylotype diversity and probably represents a microbial community with an increased specialization in cellulose degradation. Based on this conclusion, other comparisons were made only between consortia OS4, PS4, and BL6.

The comparison of the consortia from sawdust and leaf litter (PS4 and BL6) with those from straw (OS2 and OS4), revealed differences at higher taxonomic levels. Consortia OS4 and PS4 shared 29.6% of the total read count (Figure S7), OS4 and BL6—33.3% (Figure S8), PS4 and BL6—16.4% (Figure S9). In the PS4 consortium, the main changes occurred within Alphaproteobacteria. The representation of Rhizobiales was increased, but at the same time, the representation of the *Devosia* genus decreased. The presence of Burkholderiales decreased, but Alcaligenaceae (*Achromobacter*, *Pigmentiphaga*) increased. In the order Caulobacterales, the composition of phylotypes within the genus *Asticcacaulis* changed. Among other phyla, a decrease in the diversity of Verrucomicrobiota with an increase in phylotypes of the genus *Opitutus* was noted.

The BL6 consortium from leaf litter differed most from the others in the presence of a large number of representatives of the Firmicutes phylum, which were underrepresented in other consortia. The Alphaproteobacteria group was characterized by the increase in *Magnetospirillum* and the decrease in *Roseomonas* phylotypes, changes in the composition of Caulobacteraceae (phylotypes of *Caulobacter*, *Brevundimonas*, *Phenylobacterium*), decrease in *Bosea* phylotypes, and increase in *Afipia*. Gammaproteobacteria were characterized by a decrease in Comamonadaceae (only the genus *Ramlibacter* is represented in the leaf litter) with the presence of genera *Methyloversatilis* and *Achromobacter*. Within the phylum Bacteroidota, the representation of the genus *Mucilaginibacter* increased, which replaces other genera from Sphingobacteriales (*Parapedobacter*, *Olivibacter*).

For taxonomical comparisons of consortia, we used two approaches - relative abundance data and compositional analysis of balances, which allowed us to detect differences at and below genus level. Our data shows that consortia isolated from enrichment cultures of different substrates are mostly bacterial and differ in the composition of major phylotypes. Each substrate left an imprint on the overall taxonomic composition; however, the composting facilitated growth of specific microbiota, which were classified into different phylotypes, but attributed to common taxonomic groups.

2.2. Functional Analysis of Cellulolytic Consortia

Cellulolytic capacities of obtained consortia were accessed by full metagenome sequencing on the ONT platform. Four metagenomes were assembled with comparable metrics (Table 2). The number of contigs varied between 1466–1789 with the total length ranging between 86–108 Gb. Consortium BL6 had the highest total length of contigs and N50 value. Consortium OS2 had the highest number of contigs with the lowest total length.

Table 2. Characteristics of assembled metagenomes of isolated microbial consortia.

Consortium ID	Contigs	Longest Contig, bp	Total Contigs Length	N50
OS2	1879	1,582,432	86,161,676	72,582
OS4	1612	5,307,558	96,558,959	131,290
PS4	1466	2,705,151	87,318,114	112,897
BL6	1748	3,656,276	108,520,437	186,463

2.2.1. Metagenome Taxonomy

The taxonomy obtained by Kraken2 coincided with the results of the 16S rRNA gene libraries. Relative abundance could not be compared because Kraken normalizes by the number of contigs, which is very variable in the assembled genomes. Nonetheless, top taxa from consortia were detected both by amplicon libraries and metagenomes (Supplement File), including representatives of Proteobacteria, Actinobacteriota, Firmicutes, Bacteroidota, Verrucomicrobiota, Myxococcota and Planctomycetota. Some of the major representatives were annotated on the species levels: *Aquamicrobium sediminum*, *Shinella granulii*, *Devosia elaelis*, *Asticcacaulis tiandongensis*, *Opitutus terrae*, *Cellulomonas gelida* in OS2 and OS4, *Sporocytophaga myxococcoides* in OS2 and PS4, *Pseudoxanthomonas japonensis* in OS4 and BL6, *Achromobacter denitrificans*, *Ruminiclostridium hungatei* and *Methyloversatilis discipulorum* in BL6, *Youhaiella tibetensis* in PS4.

2.2.2. MAGs

Based on the metagenomic assemblies, 19 (2 from OS2, 5 from OS4, 5 from PS4-5, 6 from BL6) metagenome-assembled genomes (MAGs) were described that meet the following requirements: completeness of more than 90%, contamination of less than 5%. (Table S1). Around half of the MAGs attributed to the most abundant microorganisms, revealed by amplicon sequencing, such as *Cellulomonas*, *Sporocytophaga*, *Pseudoxanthomonas*, *Devosia*, *Asticcacaulis*, *Opitutus*, which had a great variety of pathways for lignocellulose degradation (Figure 5). There were also MAGs attributed to genera, which presented a smaller component of the consortia: *Dokdonella*, *Luteimonas*, *Afipia*, *Terricaulis*, *Methyloversatilis* and *Magnetospirillum*, which did not have as many genes with CAZy domains. Additionally, these MAGs had genes responsible for microbial interactions: quorum sensing, antimicrobial resistance, and CRISPR-Cas systems. All MAGs show genes associated with the nitrogen, sulfur, and methane metabolic activities, especially those from BL6 consortium.

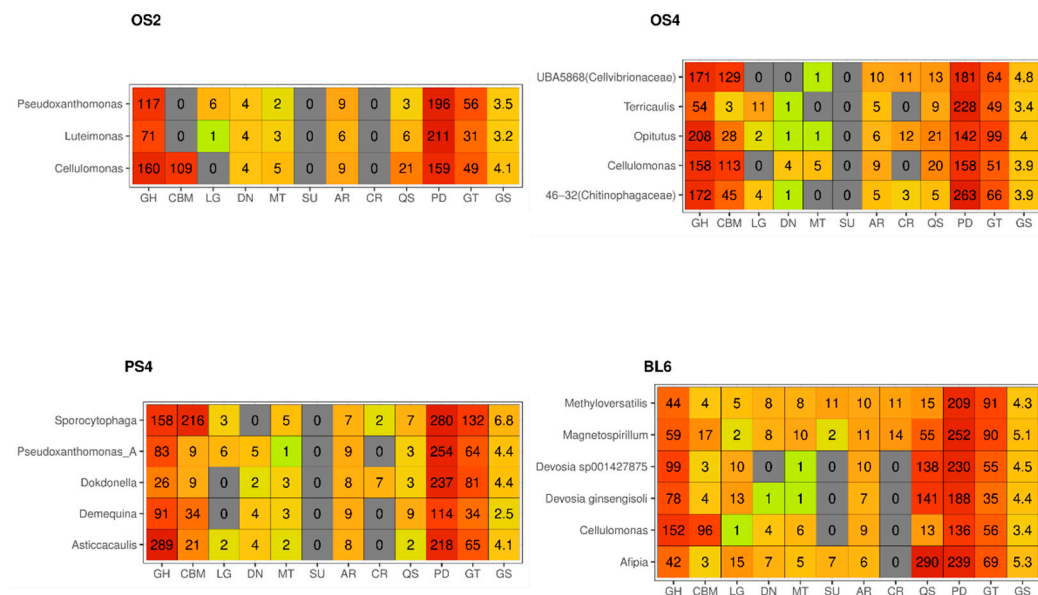


Figure 5. MAGs from the four consortia and representation of functional genes related to substrate decomposition and microbial interactions. Heatmaps are based on the DRAM annotation. Increased hit value relative to the dbCAN/EggNOG annotation caused by using a greedier algorithm. GH—Glycoside Hydrolyses (DRAM); CBM—Carbohydrate-Binding Modules (DRAM); LG—Ligninases (eLignin Database pathways, Metacyc, EC); DN—Denitrification (DRAM); MT—C1-methane (DRAM); SU—Sulfur metabolism-sulfite reductase, thiosulfate reductase, thiosulfate oxidation by SOX complex, thiosulfate \geq sulfate (DRAM); AR—Antibiotic Resistance (DRAM); CR—CRISPR (DRAM); QS—quorum sensing (KEGGs from ko02024 path); PD—Peptidases (DRAM); GT—GlycosylTransferases (DRAM); GS—Genome Size in Mb.

2.2.3. CAZy Composition

All CAZy categories were distributed similarly across all consortia (Table 3). At the first place were GHs (41.9–47%), then GTs (26.1–30.6%), CBMs (10.2–12%), CEs (9.5–11.1%), AAs (2.6–4.1%), and PLs (1.4–2.6%). Most CAZymes are modular with genes encoding several domains from one or more of the CAZy categories [37]. Thus, we tried to address the composition of CAZy modules in the metagenomes. In our consortia, the most frequently occurring pairs of modules in CAZyme genes belonged to combinations of GH, CBM and GT modules: GH13 + CBM48 (43), GH5 + GH9 (48), GH43 + GH51 (13), GT2 + GT4 (12) (Figure S10).

Table 3. CAZy module composition in metagenomes of isolated microbial consortia, % of CAZy count. AAs—Auxiliary Activities, CBMs—Carbohydrate-Binding Modules, CEs—Carbohydrate Esterases, GHs—Glycoside Hydrolases, GTs—Glycosyl Transferases, PLs—Polysaccharide Lyases.

Consortium ID	AAs	CBMs	CEs	GHs	GTs	PLs
OS2	4.1	10.8	10.7	43.1	29.2	2.2
OS4	3.7	11.3	11.1	41.9	30.6	1.4
PS4	2.6	10.2	11.1	44.3	30.0	1.8
BL6	2.7	12.0	9.5	47.0	26.1	2.6

It has been reported that CAZyme genes, along with transcription factors and transporters, are associated in clusters of several closely spaced genes, the so-called CAZyme gene clusters—CGCs [38]. CGC-finder demonstrated that, in all four consortia, around 32.5% of CAZyme genes were grouped into CGCs, which contained from 2 to 35 genes, with a median of 5 genes. BL6 consortium had the biggest cluster size and proportion of CAZyme genes in it. Most frequently CGCs were found in Rhizobieaceae (*Devosia*, *Bosea*, *Shinella*, *Bradyrhizobium*, *Agrobacterium*, *Aquamicrobium*), Xanthomonadaceae (*Pseudoxanthomonas* and *Luteimonas*), and Paenibacillaceae (*Cohnella*) (Table S3).

In total, we detected 2854 GH genes, which were attributed to approximately 120 different bacterial genera (Table S2). The highest number of GH genes were found in *Devosia* (Alphaproteobacteria), *Cohnella* (Firmicutes), *Cellulomonas* (Actinobacteriota), *Asticcacaulis* (Alphaproteobacteria), *Pseudoxanthomonas* (Gammaproteobacteria), and *Sporocytophaga* (Bacteroidota), which were detected as major genera by the amplicon sequencing. Most of the other major phylotypes identified at the amplicon sequencing stage were confirmed to have GHs, apart from *Caenimonas* and *Chryseobacterium*, which did not have GHs at all, and *Flavobacterium*, *Pirellula*, *Terrimonas*, *Dyadobacter*, and *Protochlamydia*, which had singular representatives of GH.

The most widely represented groups of GH families are shown in Figure 6. The functional pattern is similar for all consortia, with the prevailing presence of GH13, GH43, GH23, and GH3 families. However, the opposite situation could be described for CBM—their pattern differs between consortia (Figure S11). Families CBM6, CBM 35, and CBM44 were more characteristic of the consortium PS4, families CBM2 and CBM13, of OS2 and OS4, and families CBM16, CBM32, CBM61, and CBM66—of BL6. The highest proportion value of CBMs in the metagenomes was in BL6 consortium (3.2%), and the lowest in OS2 (1.7%).

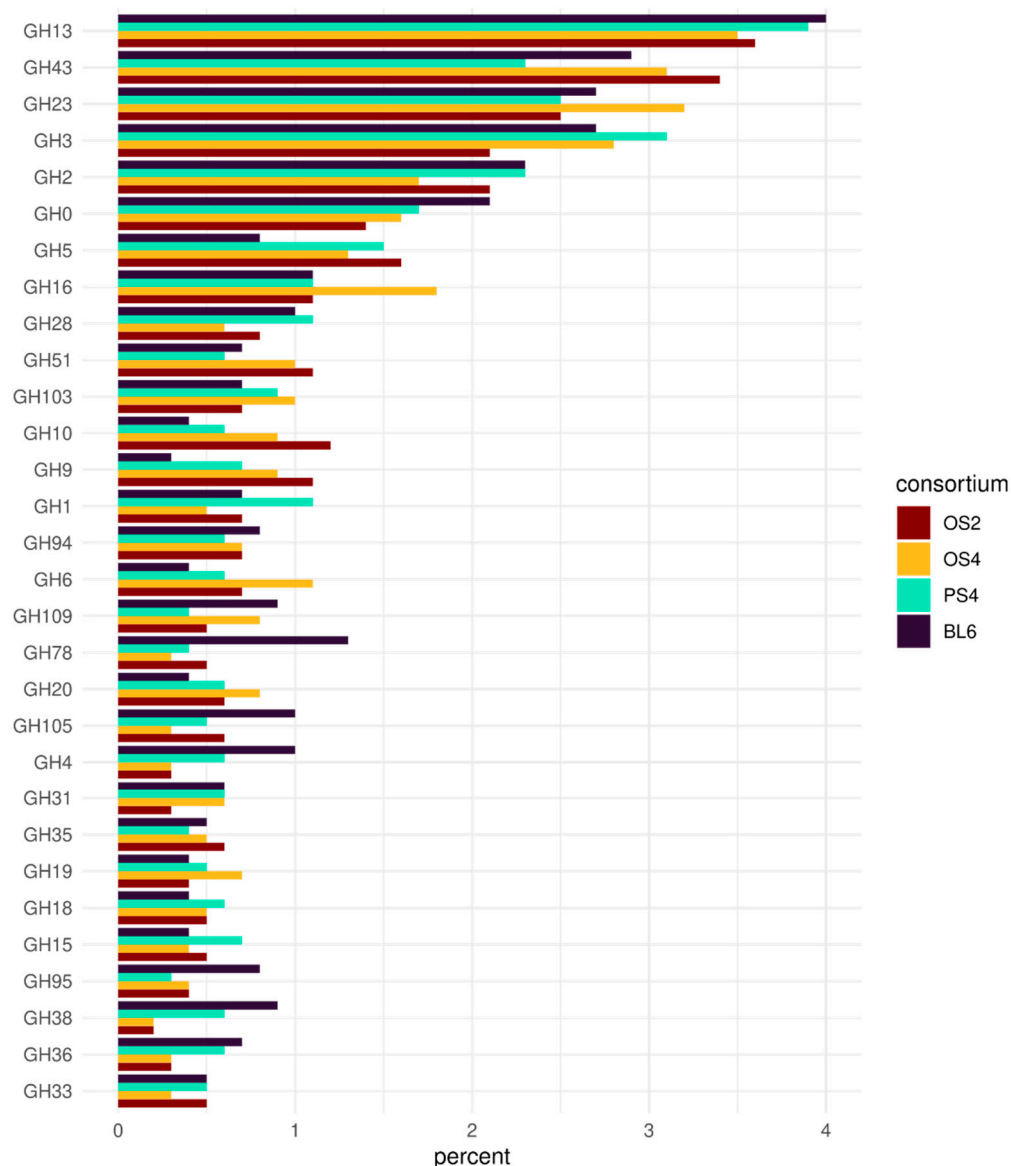


Figure 6. GH families in the four cellulolytic consortia (OS2, OS4, PS4, BL6).

Combined 16S and metagenomic data show that most of the major phylotypes are attributed to the potential active cellulolytic bacteria in the isolated microbial consortia.

2.2.4. Gene Set Enrichment Analysis

Gene Set Enrichment Analysis (GSEA) is modeled to be used with a differential expression analysis tool in order to determine whether a predefined set of genes (such as GO term or KEGG pathway) shows statistically significant differences between two biological states. Due to the specifics of our data, we applied GSEA to determine which KEGG categories, not limited to CAZy, are over-represented in metagenomes of our cellulolytic consortia. It was performed pairwise between OS4, PS4, and BL6. After the comparisons, the most characteristic KEGGs of each consortium were selected (Table S4).

The OS4 consortium was very similar to OS2 and was characterized by the over-representation of ion transport genes (TonB receptor K02014, siderophore transport K02014, K02031), ABC transporters associated with oligosaccharide transport (K02035, K02031), peptidase activity (K00281, K01256, K01354, K00681, K01941), 4-hydroxyphenylpyruvate dioxygenase (K0047), transposase (K07497), exporters (K06147, K18138, K03296), aldehyde dehydrogenase (K00128). The adhesin transporter genes (K21449, K02031) should also

be noted, which can be associated with quorum sensing and the high representation of Myxococcota in this consortium.

In the PS4 consortium we detected a higher representation of ABC transporters and permeases, often associated with oligosaccharide transport (K01990, K01992, K02004, K02025, K01998, K01998), ABC exporters (K03296, K03296), ion transport genes (TonB receptor K02014, siderophore transport K02014, heavy metals-K15726, K03455). At the same time, there was an over-representation of genes of energy metabolism for the final stages of fermentation (alcohol dehydrogenase K00001) and respiration (K02274, K01681), genes associated with translation (K01872, K01426, K01876, K01870, K03977), genes associated with cell division (K03654, K02621, K03168, K03654), histidine kinase associated with chemotaxis and cell–cell signaling (K06596, K05874, K07716), serine-threonine protein kinases (K08884, K12132). From the CAZyme genes, GH5, GH9, and GH3 were overrepresented. It is worth noting the presence of genes of a heavy metal export system. To conclude, this consortium is characterized by an over-representation of translation and cell division genes.

Consortium BL6 was enriched by the following categories: oligosaccharide transporters (K02026, K17318, K10548, K08483, K02027, K10119), including permease (K10118, K17319, K02004), ABC exporters (K01990, K06147, K18138), ion transport genes (TonB receptor K02014, siderophore transport-K02015, K02016, K07165), repair polymerases and endonucleases (K02337, K16898, K02337), genes responsible for chemotaxis (K03406, K03406), serine histidine kinase (K03407), arabinose operon regulator (K07720), sporulation regulator (K06297). All in all, in BL6 consortium, the genes for signal transduction proteins, and specific oligosaccharide transporters were overrepresented. There appears to be a definite response to nutrient deficiencies.

According to the GSEA, the main differences in cellulolytic consortia did not show in the CAZyme genes, but mostly in genes of transport and energy systems, and cell division.

2.2.5. Taxonomy Evenness in KEGGs

Classification of the four consortia on the phylum level based on ORFs, GH, CBM, GT, and CGCs revealed taxonomic unevenness between these categories (Figure 7). Proteobacteria phylum was dominant in all four consortia by ORFs, but other phyla have a proportionally higher impact of GHs and CBMs in the metagenomes in comparison to ORFs. In OS2 consortium, they were Actinobacteriota, Bacteroidota, and Verrucomicrobiota. In OS4—Myxococcota, in PS4—Actinobacteriota and Bacteroidota, in BL6—Firmicutes.

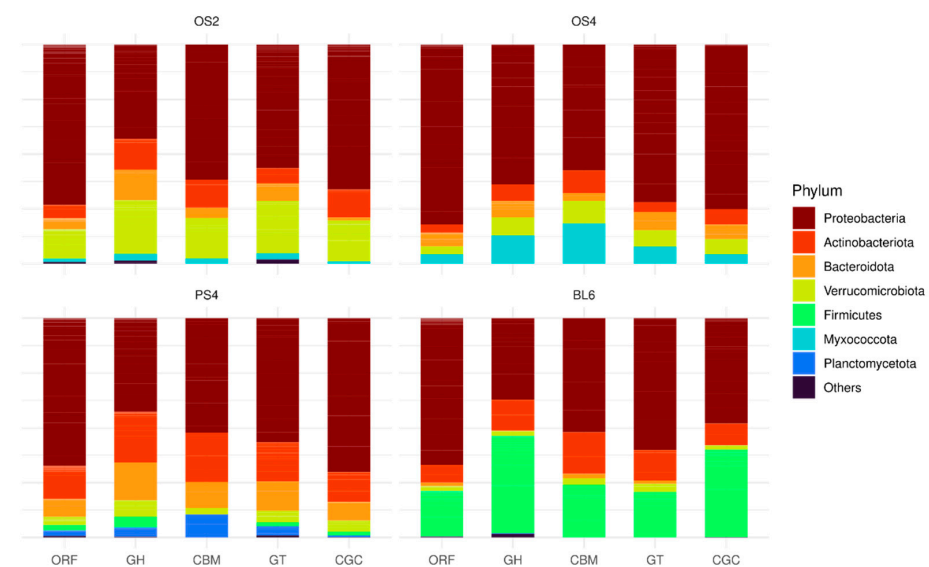


Figure 7. Taxonomic composition on the phylum level in the four cellulolytic consortia (OS2, OS4, PS4, BL6), based on ORFs, GHs, CBMs, GTs, and CGCs annotations.

To look into functional unevenness between consortia on the deeper taxonomic level, we used the inverted Simpson index for KEGG categories on the genus level. KEGGs were filtered by the number of occurrences (>50 in all metagenomes), and only whose values of the inverted Simpson index fell into the 5th percentile were left, 47 KEGGs in total (Table S5). This cut-off ensured that we were left only with large groups of KEGG categories, which were dispersed unevenly between genera (Figure S12). There was a statistically significant association (generalized linear model with quasi-Poisson family distribution) between the inverted Simpson index and the coefficient of variance of these KEGGs between consortia (Figure S13). Thus, using this approach, we elected KEGGs from different genera from different consortia. Selected KEGGs were the most often associated with the genera *Devosia*, *Cohnella*, *Shinella*, and *Bosea*. Functionally, the genes from these KEGGs are highly specific, and often overlap with those found by GSEA. The smallest evenness is characteristic of ABC transporters and membrane receptors of two-component systems. Of the enzymes associated with the metabolism of carbohydrates, it is worth highlighting the aldouronate-utilization operon components characteristic of the Paenibacillaceae family, including the regulators of oligosaccharide metabolism YesN/YesM (K07718, K07720, K03534), permeases of the aldouronate transport system (K17319, K17318, K17320). In addition, coinciding with GSEA, inverted Simpson index selected groups of oligosaccharide transporters (K10111, K10440, K01473, K15771) and exporters, most often characterized as transporters of metal ions and peptides (K09969, K10823, K09969). Among these enzymes, some groups were associated not with the metabolism of carbohydrates, but rather with the formation of a microbial community, such as quorum sensing genes (K02031, K02033). It should also be noted that CAZymes were not selected by this approach, which, together with the results of the analysis of MAGs and GSEA, allows us to say that these enzymes are not taxon-specific for the cellulolytic consortia that we obtained. Characteristically, these functional groups (ABC transporters and two-component systems), revealed by invSimpson analysis and GSEA, overlap with the functional groups associated with CAZy in the CGCs.

3. Discussion

3.1. Diversity of Members of the Cellulolytic Consortia

There is evidence that only a few microorganisms are sufficient for the destruction of lignocellulose [39]. However, the diversity of microorganisms in selected consortia from our experiment was, according to the results of the amplicon sequencing, about a hundred phylotypes. This can be explained by both methodological and biological factors. In our experiment we aimed to dilute composting lignocellulose substrate to the extent that only active cellulolytic organisms would be able to survive on a medium with filter paper as a sole carbon source. Indeed, major phylotypes from all consortia we selected attributed to genera, most commonly characterized as mesophilic aerobic and facultative anaerobic cellulose degraders from soil and sediments: *Sporocytophaga myxococcoides* [40,41], *Cellulomonas* [42–45], *Flavobacterium* [46,47], *Pseudoxanthomonas* [48,49], *Asticcacaulis* [50,51], *Paenibacillus* [52,53], and *Devosia* [54]. However, each consortium contained specific microorganisms, most likely not associated with the direct decomposition of the lignocellulosic complex. *Caenimonas*, a soil mesophile aerobe [55,56], was one of the components of OS4 consortia, but it did not have GHs according to metagenome data. Another striking example of such a component was *Bdellovibrionota*-predatory microorganism [57], found as a minor component in all consortia. Interestingly, Myxococcota, which was represented in the OS4 consortium, had a lot of GHs, but they are also characterized as predatory [58]. This could explain their prevalence in the consortium selected from the oat straw, composted for a longer period. The high diversity of the bacterial component can also be associated with the almost complete absence of eukaryotes in the dataset [59]. Despite some representation of the fungal community in amplicon ITS sequencing, most of the data were not identified in known databases. Three representatives belong to Ciliophora, which also has bacterivorous species [60]; thus, they also add to the non-cellulolytic part of the selected microbial consortia.

The relationship between microorganisms in lignocellulolytic communities can be expressed not only through competition for the substrate. As an example, lignocellulosic complexes in the process of degradation can themselves exhibit different antimicrobial activity due to decomposition into specific secondary metabolites that are toxic to a part of the microflora [61]. In addition, during cellulose decomposition, biofilm formation often occurs, especially under conditions of reduced aeration [62]. According to the amplicon sequencing, we have shown that the leaf litter consortium BL6 contains both aerobes (*Achromobacter*) [63], microaerophiles (*Methyloversatilis*, *Magnetospirillum*) [64,65], and strict anaerobes (*Ruminiclostridium*, *Anaerocolumna*, *Herbinix*) [66–68], which can be a sign of biofilm formation in this consortium, since they are often formed in nature during the breakdown of lignocellulose [69]. In the consortium OS4, we observed *Flavobacterium* as the major component, which also was reported to be able to form biofilms [70].

Taxonomic differences between consortia were revealed both in the high and low taxonomic resolution. On the phylum level, all consortia consisted predominantly of Proteobacteria, Actinobacteriota, and Bacteroidota, but each consortium had an individual imprint of minor phyla. For OS2, it was Verrucomicrobia, OS4–Myxococgota, Planctomycetota, and Firmicutes. Analysis of taxonomical composition by PhILR package showed that substrate specificity is associated with transitions at the taxonomic level below the genus. What this means is that different microbial communities from different natural substrates demonstrated the process of convergence, where similar, but not identical bacterial genera were given growth advantage. These results are enforced by metagenome data, where we can look more deeply into differences at the species level.

3.2. Functional Metagenome of the Cellulolytic Consortia

Four isolated consortia had similar profiles of GHs. Most of the GH families detected in the consortia were associated with lignocellulose decomposition, with some exceptions. The most abundant GH group in the consortia belonged to the family GH13, which predominantly consists of enzymes acting on α -glycosidic bonds, thus not partaking in lignocellulose composition [71]. Another example of a major GH group not participating in cellulose degradation is family GH23, consisting predominantly of lysozymes and peptidoglycan lyases [72], probably contributing to interaction with other members of the consortium. However, there is evidence that this family contains some unconventional genes, involved in cellulose decomposition [73]. The most abundant cellulose-related families detected in consortia (GH43, GH3, GH2, GH16, and GH1), were connected to genes encoding enzymes from the last stage of cellulose and hemicellulose decomposition: glucosidases, xylosidases, galactosidases, etc. [74]. GH families, encoding enzymes from the first stages of cellulose and hemicellulose decomposition (GH5, GH51, GH10, GH9 and GH6), were present, but less abundant [75]. Thus, isolated consortia demonstrated the later stages of lignocellulolytic activity alongside defense functions and utilization of simple sugars.

Despite the high number of CAZy modules in the individual genomes and the metagenomes as a whole, their quantities did not differ significantly in the consortia, but actual differences were detected for the proteins of the transport systems and membrane regulatory domains. Genes, characteristic of a certain consortium, were transport proteins, genes associated with two-component systems (receptors) and housekeeping genes (replication, repair, chaperones). At the same time, it is difficult to interpret what role different transport proteins play. Most often, we cannot divide this group functionally because it contains both ABC transporters associated with ion transport and peptide export. It is known that the presence of copper transport, which is a limiting macronutrient for microbial communities, from the cell is associated with ligninase activity (aldehyde dehydrogenase), [76]. It should not be ruled out that ABC transporters are an important part of protective enzymes, which in turn can be associated with the important role of microbial selection in the community [77]. It should be noted that the presence of exporters can be associated both with the export of catalases and with the protective function associated with microbial–microbial interactions. This is also supported by the fact that all metagenomes of

the consortia have antibiotic resistance genes (kanamycin, sulfisoxazole), CRISPR systems, receptors associated with quorum sensing and toxin/antitoxin systems. Of the genes that may play a role in the formation of biofilms in the studied communities, adhesins and spermidine synthesis genes are worth noting, which is in good agreement with the data obtained based on amplicon sequencing data [78]. Among the transport systems, specific cassettes associated with the transport of oligosaccharides were found. This agrees with the fact that around half of detected CGCs contained transport proteins.

In our consortia, we detected widely known cellulolytic bacteria, such as *Sporocytophaga* and *Cellulomonas*. At the same time, we detected the genomes of cellulolytic microorganisms, which are not widely represented in the literature devoted to biotechnology, as a major component of the consortia. For example, representatives of Verrucomicrobia are very widely represented in the soil, but many of them are uncultivated [79]. The genome of the representative of Verrucomicrobia (*Opitutus*) collected by us contained a high representation of glycoside hydrolases. *Magnetospirillum* isolated from the leaf litter consortium is also an unexpected representative of cellulolytic communities. In addition, the presence of microorganisms associated with methylotrophy, *Methyloversatilis discipulorum* [80], was shown in the leaf litter consortium. In the article devoted to the early decomposition of oak leaves, the presence of the dominance of *Methylosinus trichosporium*, which also has methylotrophic activity, was shown in the early stages of decomposition [80]. *Magnetospirillum* has a similar metabolic profile [81]. Methylotrophy is a specific feature of the microbial process of lignocellulose degradation, in contrast to the fungal one. The methylotrophic component found in the litter consortium is of considerable interest to biotechnologies because it allows you to increase the yield of an organic product [82].

4. Materials and Methods

Oat straw, pine sawdust, and birch leaf litter were used to isolate minimal cellulolytic consortia. The experiment was laid out in October 2018. Three replicate samples of each substrate weighing 60 g were moistened up to 60% of the total capacity and mixed with mineral fertilizer (NPK elements in 16:16:16 ratio) to a concentration of 10 mg/g of the substrate. All substrates and fertilizer were non-sterile and did not have any additional microbial inoculum. Fertilizer was added to facilitate decomposition by removing macronutrient deficiency, characteristic of C-rich biomass. The mixture of each substrate in three replicates was composted in plastic containers at 28 °C for 6 months. At two, four, and six months it was remixed, and each substrate was sampled for the enrichment test. Each sample was processed with a series of 10-fold dilutions in sterile water, which were inoculated into 5 mL of Hutchinson liquid medium (g/L: K_2HPO_4 —1; NaCl—0.1; $CaCl_2$ —0.1; $FeCl_3$ —0.01; $MgSO_4 \cdot 7H_2O$ —0.3; $NaNO_3$ —2.5) with ashless filter paper as the only carbon source [83]. We looked for microbial consortia grown from the largest dilutions, which preserved the ability to macerate the filter paper. After the experiment, only four stable microbial consortia with reproducible phenotypes and cellulolytic activity after several months of transfers were selected for the 16S rRNA libraries and full metagenome sequencing.

For DNA extraction fresh microbial consortia were incubated in tubes for two weeks in a new portion of Hutchinson liquid medium. In total, six replicate cultures for each of the four consortia were acquired. After the incubation period, the contents of the tubes were centrifuged, and the supernatant was removed. The cell debris with the remaining filter paper was resuspended in SL1 buffer from NucleoSpin® Soil Kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany) and ground using a Precellys 24 homogenizer (Bertin Technologies, Saint-Quentin en Yveline, France). This mixture was used for DNA extraction according to the manufacturer's recommendations. Construction and sequencing of the 16S rRNA amplicon libraries were performed on the Illumina MiSeq platform (Illumina, Inc., San Diego, CA, USA) as described previously [84].

Amplicon libraries were processed using the DADA2 pipeline [85] in the R software environment [86]. Taxonomic identification was carried out using the Silva 138 database [87] for 16S rRNA gene sequences and the Unite database [88] for ITS sequences. The phylo-

genetic tree was constructed using the IQ-TREE 2.1.2 program [89]. Further processing was carried out using the phyloseq [90] and ampvis2 [91] packages. Alpha diversity was accessed by three indexes: Observed, Shannon [92], and inverted Simpson [93], with the significance of mean differences between them calculated by the Mann–Whitney test [94]. Beta diversity was accessed by PCoA [95] with the Bray–Curtis distance matrix [96]. A composite data analysis (phylogenetic isometric log-ratio) based on the PhILR package [97] was used to search for distinct groups of microorganisms.

For the metagenome sequencing, all six DNA samples were mixed for each consortium. The resulting four mixtures were used for the library construction with the ligation sequencing kit (SQK-LSK110) using genomic DNA by ligation protocol (SQK-LSK110) and sequencing on the MinION platform with the Flow Cell R9.4.1 (Oxford Nanopore Technologies, Oxford, UK). The initial processing of the full metagenomic reads was conducted with the Guppy 3.6.1 neural network model [98]. The assembly was performed in the metaFlye 2.8 program [99], additional cleaning of the resulting contigs was carried out using algorithms included in Racon [100] and medaka [101].

Taxonomic classification of contig-assembled metagenome sequence data was performed by Kraken2 [102]. Prokaryotic contigs were assigned using the pre-built database based on GTDB 202 [103], eukaryotic on the pre-built custom database maxikraken2 1903 based on RefSeq NCBI [104]. Visualization of relative abundance was performed in Krona [105].

The annotation of Carbohydrate-Active Enzymes in the assembled contigs was obtained using dbCAN [106]. An ORF match was considered if two tools from DIAMOND [107], HMMER or Hotpep [108] annotated it. The search of CAZyme gene clusters (CGCs) in assembled contigs was performed by the CGC-Finder [109]. The rest of the genes were annotated using EggNOG-mapper (v.2.1.6) pipeline in ultra-sensitive DIAMOND mode with the `-framefix` flag [110,111]. Gene overrepresentation analysis (GSEA) was performed using the ClusterProfiler (v 1.3.1) package [112] using EggNOG-mapper with KEGG and COG functional classifications with Fisher test [113] and FDR correction method [114] for multiple comparisons. Taxonomic unevenness of KEGGs was calculated in vegan [115].

High-quality MAGs were isolated from the obtained assemblies using the MetaBat2 program [116] using the following parameters: over 90% completeness and less than 5% heterogeneity as defined by the CheckM program [117]. MAGs were taxonomically annotated with the GTDB-Tk toolkit using the HMMER, pplacer, FastANI, and FastTree algorithms based on the GTDB 202 database [118]. Functional annotation was carried out using the DRAM pipeline using the UniRef90, PFAM, dbCAN, and MEROPS databases [119]. The code is available at https://github.com/crabron/cellulolytic_consortia, accessed on 12 August 2022.

5. Conclusions

In our work we isolated four stable cellulolytic consortia from various natural ligno-cellulose substrates by making enrichment cultures from diluted compost with the filter paper as a sole carbon source. Each consortium was a complex microbial community with distinctive phenotypic features. At the core of every consortium were cellulolytic bacteria from various genera; however, in addition to them, auxiliary components were present, such as biofilm inhabitants, predators, or methylotrophs. Being taxonomically different, the composition of CAzyme genes in the consortia coincided. The main differences between consortia were detected in the taxon-specific transport and regulatory genes. The similar conditions of selection and cultivation functionally brought together taxonomically different consortia; however, each substrate left an individual imprint on their composition.

Supplementary Materials: The supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms231810779/s1>.

Author Contributions: Conceptualization, G.V.G. and A.K.K.; methodology, G.V.G., O.V.O. and T.O.L.; software, G.V.G.; validation, G.V.G., A.M.A. and A.K.K.; formal analysis, G.V.G.; investigation, G.V.G. and A.K.K.; resources, A.G.P., A.A.K. and T.S.A.; data curation, G.V.G.; writing—original draft preparation, G.V.G. and A.K.K.; writing—review and editing, G.V.G., A.M.A. and A.K.K.; visualization, G.V.G. and A.K.K.; supervision, E.E.A.; project administration, E.E.A.; funding acquisition, E.E.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research and the APC were funded by the Russian Science Foundation, grant number 18-16-00073.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All acquired data is available at the BioProject Database (<https://www.ncbi.nlm.nih.gov/bioproject/>), accessed on 12 August 2022) via Accession: PRJNA844230.

Acknowledgments: We thank the Centre for Genomic Technologies, Proteomics and Cell Biology (ARRIAM, Russia) for performing the preparation and sequencing of gene and genomic libraries.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Naik, S.N.; Goud, V.V.; Rout, P.K.; Dalai, A.K. Production of First and Second Generation Biofuels: A Comprehensive Review. *Renew. Sustain. Energy Rev.* **2010**, *14*, 578–597. [\[CrossRef\]](#)
2. Howard, R.L.; Abotsi, E.; van Rensburg, E.J.; Howard, S. Lignocellulose Biotechnology: Issues of Bioconversion and Enzyme Production. *Afr. J. Biotechnol.* **2003**, *2*, 602–619. [\[CrossRef\]](#)
3. Martins, D.A.B.; do Prado, H.F.A.; Leite, R.S.R.; Ferreira, H.; de Moretti, M.M.S.; da Silva, R.; Gomes, E. *Agroindustrial Wastes as Substrates for Microbial Enzymes Production and Source of Sugar for Bioethanol Production*; IntechOpen: London, UK, 2011; pp. 319–360.
4. Abdeshahian, P.; Kadier, A.; Rai, P.K.; da Silva, S.S. Lignocellulose as a Renewable Carbon Source for Microbial Synthesis of Different Enzymes. In *Lignocellulosic Biorefining Technologies*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2020; pp. 185–202.
5. Fellows, C.M.; Brown, T.C.; Doherty, W.O.S. Lignocellulosics as a Renewable Feedstock for Chemical Industry: Chemicals from Lignin. In *Green Chemistry for Environmental Remediation*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2011; pp. 561–610.
6. Sánchez, C. Lignocellulosic Residues: Biodegradation and Bioconversion by Fungi. *Biotechnol. Adv.* **2009**, *27*, 185–194. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Weidener, D.; Dama, M.; Dietrich, S.K.; Ohrem, B.; Pauly, M.; Leitner, W.; Domínguez de María, P.; Grande, P.M.; Klose, H. Multiscale Analysis of Lignocellulose Recalcitrance towards OrganoCat Pretreatment and Fractionation. *Biotechnol. Biofuels* **2020**, *13*, 155. [\[CrossRef\]](#)
8. Pauly, M.; Keegstra, K. Cell-Wall Carbohydrates and Their Modification as a Resource for Biofuels. *Plant J.* **2008**, *54*, 559–568. [\[CrossRef\]](#)
9. Xu, F.; Li, Y. Biomass Digestion. In *Encyclopedia of Sustainable Technologies*; Abraham, M.A., Ed.; Elsevier: Oxford, UK, 2017; pp. 197–204.
10. Saake, B.; Lehnen, R. Lignin. In *Ullmann's Encyclopedia of Industrial Chemistry*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2007.
11. Horn, S.J.; Vaaje-Kolstad, G.; Westereng, B.; Eijsink, V. Novel Enzymes for the Degradation of Cellulose. *Biotechnol. Biofuels* **2012**, *5*, 45. [\[CrossRef\]](#)
12. Várnai, A.; Huikko, L.; Pere, J.; Siika-aho, M.; Viikari, L. Synergistic Action of Xylanase and Mannanase Improves the Total Hydrolysis of Softwood. *Bioresour. Technol.* **2011**, *102*, 9096–9104. [\[CrossRef\]](#)
13. Bourne, Y.; Henrissat, B. Glycoside Hydrolases and Glycosyltransferases: Families and Functional Modules. *Curr. Opin. Struct. Biol.* **2001**, *11*, 593–600. [\[CrossRef\]](#)
14. Boraston, A.B.; Bolam, D.N.; Gilbert, H.J.; Davies, G.J. Carbohydrate-Binding Modules: Fine-Tuning Polysaccharide Recognition. *Biochem. J.* **2004**, *382*, 769–781. [\[CrossRef\]](#)
15. Carrard, G.; Koivula, A.; Söderlund, H.; Béguin, P. Cellulose-Binding Domains Promote Hydrolysis of Different Sites on Crystalline Cellulose. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 10342–10347. [\[CrossRef\]](#)
16. Linder, M.; Teeri, T.T. The Roles and Function of Cellulose-Binding Domains. *J. Biotechnol.* **1997**, *57*, 15–28. [\[CrossRef\]](#)
17. Williams, G.J.; Thorson, J.S. Natural Product Glycosyltransferases: Properties and Applications. In *Advances in Enzymology and Related Areas of Molecular Biology*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2008; pp. 55–119.
18. Coutinho, P.M.; Deleury, E.; Davies, G.J.; Henrissat, B. An Evolving Hierarchical Family Classification for Glycosyltransferases. *J. Mol. Biol.* **2003**, *328*, 307–317. [\[CrossRef\]](#)

19. Martínez, A.T.; Speranza, M.; Ruiz-Dueñas, F.J.; Ferreira, P.; Camarero, S.; Guillén, F.; Martínez, M.J.; Gutiérrez, A.; del Río, J.C. Biodegradation of Lignocellulosics: Microbial, Chemical, and Enzymatic Aspects of the Fungal Attack of Lignin. *Int. Microbiol.* **2005**, *8*, 195–204. [[PubMed](#)]
20. Drula, E.; Garron, M.-L.; Dogan, S.; Lombard, V.; Henrissat, B.; Terrapon, N. The Carbohydrate-Active Enzyme Database: Functions and Literature. *Nucleic Acids Res.* **2022**, *50*, D571–D577. [[CrossRef](#)]
21. Henrissat, B. A Classification of Glycosyl Hydrolases Based on Amino Acid Sequence Similarities. *Biochem. J.* **1991**, *280 Pt 2*, 309–316. [[CrossRef](#)]
22. Levasseur, A.; Drula, E.; Lombard, V.; Coutinho, P.M.; Henrissat, B. Expansion of the Enzymatic Repertoire of the CAZy Database to Integrate Auxiliary Redox Enzymes. *Biotechnol. Biofuels* **2013**, *6*, 41. [[CrossRef](#)]
23. Andlar, M.; Rezić, T.; Mardetko, N.; Kracher, D.; Ludwig, R.; Šantek, B. Lignocellulose Degradation: An Overview of Fungi and Fungal Enzymes Involved in Lignocellulose Degradation. *Eng. Life Sci.* **2018**, *18*, 768–778. [[CrossRef](#)]
24. Zang, X.; Liu, M.; Wang, H.; Fan, Y.; Zhang, H.; Liu, J.; Xing, E.; Xu, X.; Li, H. The Distribution of Active β -Glucosidase-Producing Microbial Communities in Composting. *Can. J. Microbiol.* **2017**, *63*, 998–1008. [[CrossRef](#)]
25. Chukwuma, O.B.; Rafatullah, M.; Tajarudin, H.A.; Ismail, N. A Review on Bacterial Contribution to Lignocellulose Breakdown into Useful Bio-Products. *Int. J. Environ. Res. Public Health* **2021**, *18*, 6001. [[CrossRef](#)]
26. Wang, C.; Dong, D.; Wang, H.; Müller, K.; Qin, Y.; Wang, H.; Wu, W. Metagenomic Analysis of Microbial Consortia Enriched from Compost: New Insights into the Role of Actinobacteria in Lignocellulose Decomposition. *Biotechnol. Biofuels* **2016**, *9*, 22. [[CrossRef](#)]
27. DeAngelis, K.M.; Sharma, D.; Varney, R.; Simmons, B.; Isern, N.G.; Markillie, L.M.; Nicora, C.; Norbeck, A.D.; Taylor, R.C.; Aldrich, J.T.; et al. Evidence Supporting Dissimilatory and Assimilatory Lignin Degradation in *Enterobacter lignolyticus* SCF1. *Front. Microbiol.* **2013**, *4*, 280. [[CrossRef](#)] [[PubMed](#)]
28. Lam, M.Q.; Oates, N.C.; Thevarajoo, S.; Tokiman, L.; Goh, K.M.; McQueen-Mason, S.J.; Bruce, N.C.; Chong, C.S. Genomic Analysis of a Lignocellulose Degrading Strain from the Underexplored Genus *Meridianimaribacter*. *Genomics* **2020**, *112*, 952–960. [[CrossRef](#)]
29. Ventorino, V.; Ionata, E.; Birolo, L.; Montella, S.; Marcolongo, L.; de Chiaro, A.; Espresso, F.; Faraco, V.; Pepe, O. Lignocellulose-Adapted Endo-Cellulase Producing *Streptomyces* Strains for Bioconversion of Cellulose-Based Materials. *Front. Microbiol.* **2016**, *7*, 2061. [[CrossRef](#)] [[PubMed](#)]
30. Cortes-Tolalpa, L.; Jiménez, D.J.; de Lima Brossi, M.J.; Salles, J.F.; van Elsas, J.D. Different Inocula Produce Distinctive Microbial Consortia with Similar Lignocellulose Degradation Capacity. *Appl. Microbiol. Biotechnol.* **2016**, *100*, 7713–7725. [[CrossRef](#)]
31. Cortes-Tolalpa, L.; Wang, Y.; Salles, J.F.; van Elsas, J.D. Comparative Genome Analysis of the Lignocellulose Degrading Bacteria *Citrobacter freundii* So4 and *Sphingobacterium multivorum* W15. *Front. Microbiol.* **2020**, *11*, 1–15. [[CrossRef](#)] [[PubMed](#)]
32. Wongwilaiwalin, S.; Rattanachomsri, U.; Laothanachareon, T.; Eurwilaichitr, L.; Igarashi, Y.; Champreda, V. Analysis of a Thermophilic Lignocellulose Degrading Microbial Consortium and Multi-Species Lignocellulolytic Enzyme System. *Enzyme Microb. Technol.* **2010**, *47*, 283–290. [[CrossRef](#)]
33. Zuroff, T.R.; Xiques, S.B.; Curtis, W.R. Consortia-Mediated Bioprocessing of Cellulose to Ethanol with a Symbiotic *Clostridium* Phytofermentans/Yeast Co-Culture. *Biotechnol. Biofuels* **2013**, *6*, 59. [[CrossRef](#)]
34. Cheng, Y.F.; Edwards, J.E.; Allison, G.G.; Zhu, W.-Y.; Theodorou, M.K. Diversity and Activity of Enriched Ruminant Cultures of Anaerobic Fungi and Methanogens Grown Together on Lignocellulose in Consecutive Batch Culture. *Bioresour. Technol.* **2009**, *100*, 4821–4828. [[CrossRef](#)]
35. Poszytek, K.; Cieczkowska, M.; Sklodowska, A.; Drewniak, L. Microbial Consortium with High Cellulolytic Activity (MCHCA) for Enhanced Biogas Production. *Front. Microbiol.* **2016**, *7*, 324. [[CrossRef](#)]
36. Gavande, P.V.; Basak, A.; Sen, S.; Lepcha, K.; Murmu, N.; Rai, V.; Mazumdar, D.; Saha, S.P.; Das, V.; Ghosh, S. Functional Characterization of Thermotolerant Microbial Consortium for Lignocellulolytic Enzymes with Central Role of Firmicutes in Rice Straw Depolymerization. *Sci. Rep.* **2021**, *11*, 3032. [[CrossRef](#)]
37. Medie, F.M.; Davies, G.J.; Drancourt, M.; Henrissat, B. Genome Analyses Highlight the Different Biological Roles of Cellulases. *Nat. Rev. Microbiol.* **2012**, *10*, 227–234. [[CrossRef](#)]
38. Zhang, H.; Yohe, T.; Huang, L.; Entwistle, S.; Wu, P.; Yang, Z.; Busk, P.K.; Xu, Y.; Yin, Y. DbCAN2: A Meta Server for Automated Carbohydrate-Active Enzyme Annotation. *Nucleic Acids Res.* **2018**, *46*, W95–W101. [[CrossRef](#)]
39. Puentes-Téllez, P.E.; Falcao Salles, J. Construction of Effective Minimal Active Microbial Consortia for Lignocellulose Degradation. *Microb. Ecol.* **2018**, *76*, 419–429. [[CrossRef](#)]
40. Liu, L.; Gao, P.; Chen, G.; Wang, L. Draft Genome Sequence of Cellulose-Digesting Bacterium *Sporocytophaga myxococcoides* PG-01. *Genome Announc.* **2014**, *2*, e01154-14. [[CrossRef](#)]
41. Taillefer, M.; Arntzen, M.Ø.; Henrissat, B.; Pope, P.B.; Larsbrink, J. Proteomic Dissection of the Cellulolytic Machineries Used by Soil-Dwelling Bacteroidetes. *mSystems* **2018**, *3*, e00240-18. [[CrossRef](#)]
42. Elberson, M.A.; Malekzadeh, F.; Yazdi, M.T.; Kameranpour, N.; Noori-Dalooi, M.R.; Matte, M.H.; Shahamat, M.; Colwell, R.R.; Sowers, K.R. *Cellulomonas persica* sp. nov. and *Cellulomonas iranensis* sp. nov., Mesophilic Cellulose-Degrading Bacteria Isolated from Forest Soils. *Int. J. Syst. Evol. Microbiol.* **2000**, *50 Pt 3*, 993–996. [[CrossRef](#)] [[PubMed](#)]
43. Poulsen, H.V.; Willink, F.W.; Ingvorsen, K. Aerobic and Anaerobic Cellulase Production by *Cellulomonas Uda*. *Arch. Microbiol.* **2016**, *198*, 725–735. [[CrossRef](#)]

44. Takeuchi, Y.; Khawdas, W.; Aso, Y.; Ohara, H. Microbial Fuel Cells Using *Cellulomonas* spp. with Cellulose as Fuel. *J. Biosci. Bioeng.* **2017**, *123*, 358–363. [[CrossRef](#)]
45. Vladut-Talor, M.; Kauri, T.; Kushner, D.J. Effects of Cellulose on Growth, Enzyme Production, and Ultrastructure of a *Cellulomonas* Species. *Arch. Microbiol.* **1986**, *144*, 191–195. [[CrossRef](#)]
46. Herrera, L.M.; Braña, V.; Franco Fraguas, L.; Castro-Sowinski, S. Characterization of the Cellulase-Secretome Produced by the Antarctic Bacterium *Flavobacterium* sp. AUG42. *Microbiol. Res.* **2019**, *223–225*, 13–21. [[CrossRef](#)]
47. Kim, H.; Yu, S.M. *Flavobacterium nackdongense* sp. nov., a Cellulose-Degrading Bacterium Isolated from Sediment. *Arch. Microbiol.* **2020**, *202*, 591–595. [[CrossRef](#)] [[PubMed](#)]
48. Hou, L.; Jiang, J.; Xu, Z.; Zhou, Y.; Leung, F.C.-C. Complete Genome Sequence of *Pseudoxanthomonas suwonensis* Strain J1, a Cellulose-Degrading Bacterium Isolated from Leaf- and Wood-Enriched Soil. *Genome Announc.* **2015**, *3*, e00614-15. [[CrossRef](#)] [[PubMed](#)]
49. Kumar, M.; Revathi, K.; Khanna, S. Biodegradation of Cellulosic and Lignocellulosic Waste by *Pseudoxanthomonas* sp R-28. *Carbohydr. Polym.* **2015**, *134*, 761–766. [[CrossRef](#)] [[PubMed](#)]
50. Eichorst, S.A.; Kuske, C.R. Identification of Cellulose-Responsive Bacterial and Fungal Communities in Geographically and Edaphically Different Soils by Using Stable Isotope Probing. *Appl. Environ. Microbiol.* **2012**, *78*, 2316–2327. [[CrossRef](#)]
51. Stursová, M.; Zifčáková, L.; Leigh, M.B.; Burgess, R.; Baldrian, P. Cellulose Utilization in Forest Litter and Soil: Identification of Bacterial and Fungal Decomposers. *FEMS Microbiol. Ecol.* **2012**, *80*, 735–746. [[CrossRef](#)]
52. López-Mondéjar, R.; Zühlke, D.; Větrovský, T.; Becher, D.; Riedel, K.; Baldrian, P. Decoding the Complete Arsenal for Cellulose and Hemicellulose Deconstruction in the Highly Efficient Cellulose Decomposer *Paenibacillus* O199. *Biotechnol. Biofuels* **2016**, *9*, 104. [[CrossRef](#)]
53. Kim, E.S.; Kim, B.-S.; Kim, K.-Y.; Woo, H.-M.; Lee, S.-M.; Um, Y. Aerobic and Anaerobic Cellulose Utilization by *Paenibacillus* sp. CAA11 and Enhancement of Its Cellulolytic Ability by Expressing a Heterologous Endoglucanase. *J. Biotechnol.* **2018**, *268*, 21–27. [[CrossRef](#)]
54. Wilhelm, R.C.; Pepe-Ranne, C.; Weisenhorn, P.; Lipton, M.; Buckley, D.H. Competitive Exclusion and Metabolic Dependency among Microorganisms Structure the Cellulose Economy of an Agricultural Soil. *mBio* **2021**, *12*, e03099-20. [[CrossRef](#)]
55. Dahal, R.H.; Lee, H.; Chaudhary, D.K.; Kim, D.-Y.; Son, J.; Kim, J.; Ka, J.-O.; Kim, D.-U. *Caenimonas soli* sp. Nov., Isolated from Soil. *Arch. Microbiol.* **2021**, *203*, 1123–1129. [[CrossRef](#)]
56. Kim, S.-J.; Weon, H.-Y.; Kim, Y.-S.; Moon, J.Y.; Seok, S.J.; Hong, S.-B.; Kwon, S.-W. *Caenimonas terrae* sp. Nov., Isolated from a Soil Sample in Korea, and Emended Description of the Genus *Caenimonas* Ryu et al. 2008. *J. Microbiol.* **2012**, *50*, 864–868. [[CrossRef](#)]
57. Johnke, J.; Fraune, S.; Bosch, T.C.G.; Hentschel, U.; Schulenburg, H. *Bdellovibrio* and Like Organisms Are Predictors of Microbiome Diversity in Distinct Host Groups. *Microb. Ecol.* **2020**, *79*, 252–257. [[CrossRef](#)] [[PubMed](#)]
58. Murphy, C.L.; Yang, R.; Decker, T.; Cavalliere, C.; Andreev, V.; Bircher, N.; Cornell, J.; Dohmen, R.; Pratt, C.J.; Grinnell, A.; et al. Genomes of Novel Myxococcota Reveal Severely Curtailed Machineries for Predation and Cellular Differentiation. *Appl. Environ. Microbiol.* **2021**, *87*, e0170621. [[CrossRef](#)] [[PubMed](#)]
59. Mille-Lindblom, C.; Tranvik, L.J. Antagonism between Bacteria and Fungi on Decomposing Aquatic Plant Litter. *Microb. Ecol.* **2003**, *45*, 173–182. [[CrossRef](#)] [[PubMed](#)]
60. Pajdak-Stós, A.; Sobczyk, M.; Fiałkowska, E.; Kocerba-Soroka, W.; Fyda, J. The Effect of Three Different Predatory Ciliate Species on Activated Sludge Microfauna. *Eur. J. Protistol.* **2017**, *58*, 87–93. [[CrossRef](#)]
61. Lee, J.A.; Baugh, A.C.; Shevalier, N.J.; Strand, B.; Stolyar, S.; Marx, C.J. Cross-Feeding of a Toxic Metabolite in a Synthetic Lignocellulose-Degrading Microbial Community. *Microorganisms* **2021**, *9*, 321. [[CrossRef](#)] [[PubMed](#)]
62. Dumitrache, A.; Wolfardt, G.M.; Allen, D.G.; Liss, S.N.; Lynd, L.R. Tracking the Cellulolytic Activity of *Clostridium Thermocellum* Biofilms. *Biotechnol. Biofuels* **2013**, *6*, 175. [[CrossRef](#)]
63. Betts, G. 23—Other Spoilage Bacteria. In *Food Spoilage Microorganisms*; de Blackburn, C.W., Ed.; Woodhead Publishing Series in Food Science, Technology and Nutrition; Woodhead Publishing: Sawston, UK, 2006; pp. 668–693.
64. Blakemore, R. Magnetotactic Bacteria. *Science* **1975**, *190*, 377–379. [[CrossRef](#)]
65. Smalley, N.E.; Taipale, S.; De Marco, P.; Doronina, N.V.; Kyrpides, N.; Shapiro, N.; Woyke, T.; Kalyuzhnaya, M.G. Functional and Genomic Diversity of Methylophilic Rhodocyclaceae: Description of *Methyloversatilis discipulorum* sp. nov. *Int. J. Syst. Evol. Microbiol.* **2015**, *65*, 2227–2233. [[CrossRef](#)]
66. Koeck, D.E.; Ludwig, W.; Wanner, G.; Zverlov, V.V.; Liebl, W.; Schwarz, W.H. *Herbinix hemicellulosilytica* gen. nov., sp. nov., a Thermophilic Cellulose-Degrading Bacterium Isolated from a Thermophilic Biogas Reactor. *Int. J. Syst. Evol. Microbiol.* **2015**, *65*, 2365–2371. [[CrossRef](#)]
67. Ueki, A.; Ohtaki, Y.; Kaku, N.; Ueki, K. Descriptions of *Anaerotaenia torta* gen. nov., sp. nov. and *Anaerocolumna cellulositytica* gen. nov., sp. nov. Isolated from a Methanogenic Reactor of Cattle Waste and Reclassification of *Clostridium aminovalericum*, *Clostridium jejuense* and *Clostridium xylanovorans* as *Anaerocolumna* Species. *Int. J. Syst. Evol. Microbiol.* **2016**, *66*, 2936–2943. [[CrossRef](#)]
68. Wu, K.; Cheng, L. Ruminiclostridium. In *Bergey's Manual of Systematics of Archaea and Bacteria*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2021; pp. 1–11.
69. Brethauer, S.; Shahab, R.L.; Studer, M.H. Impacts of Biofilms on the Conversion of Cellulose. *Appl. Microbiol. Biotechnol.* **2020**, *104*, 5201–5212. [[CrossRef](#)] [[PubMed](#)]

70. Pollet, T.; Berdjeb, L.; Garnier, C.; Durrieu, G.; Le Poupon, C.; Misson, B.; Jean-François, B. Prokaryotic Community Successions and Interactions in Marine Biofilms: The Key Role of Flavobacteriia. *FEMS Microbiol. Ecol.* **2018**, *94*, fiy083. [[CrossRef](#)]
71. Stam, M.R.; Danchin, E.G.J.; Rancurel, C.; Coutinho, P.M.; Henrissat, B. Dividing the Large Glycoside Hydrolase Family 13 into Subfamilies: Towards Improved Functional Annotations of Alpha-Amylase-Related Proteins. *Protein Eng. Des. Sel.* **2006**, *19*, 555–562. [[CrossRef](#)]
72. Brás, J.L.A.; Cartmell, A.; Carvalho, A.L.M.; Verzé, G.; Bayer, E.A.; Vazana, Y.; Correia, M.A.S.; Prates, J.A.M.; Ratnaparkhe, S.; Boraston, A.B.; et al. Structural Insights into a Unique Cellulase Fold and Mechanism of Cellulose Hydrolysis. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 5237–5242. [[CrossRef](#)] [[PubMed](#)]
73. López-Mondéjar, R.; Zühlke, D.; Becher, D.; Riedel, K.; Baldrian, P. Cellulose and Hemicellulose Decomposition by Forest Soil Bacteria Proceeds by the Action of Structurally Variable Enzymatic Systems. *Sci. Rep.* **2016**, *6*, 25279. [[CrossRef](#)]
74. Yeoman, C.J.; Han, Y.; Dodd, D.; Schroeder, C.M.; Mackie, R.I.; Cann, I.K.O. Chapter 1—Thermostable Enzymes as Biocatalysts in the Biofuel Industry. In *Advances in Applied Microbiology*; Academic Press: Cambridge, MA, USA, 2010; Volume 70, pp. 1–55.
75. Aspeborg, H.; Coutinho, P.M.; Wang, Y.; Brumer, H.; Henrissat, B. Evolution, Substrate Specificity and Subfamily Classification of Glycoside Hydrolase Family 5 (GH5). *BMC Evol. Biol.* **2012**, *12*, 186. [[CrossRef](#)] [[PubMed](#)]
76. Kamimura, N.; Goto, T.; Takahashi, K.; Kasai, D.; Otsuka, Y.; Nakamura, M.; Katayama, Y.; Fukuda, M.; Masai, E. A Bacterial Aromatic Aldehyde Dehydrogenase Critical for the Efficient Catabolism of Syringaldehyde. *Sci. Rep.* **2017**, *7*, 44422. [[CrossRef](#)]
77. Zeng, Y.; Charkowski, A.O. The Role of ATP-Binding Cassette Transporters in Bacterial Phytopathogenesis. *Phytopathology* **2021**, *111*, 600–610. [[CrossRef](#)]
78. Li, B.; Maezato, Y.; Kim, S.H.; Kurihara, S.; Liang, J.; Michael, A.J. Polyamine-Independent Growth and Biofilm Formation, and Functional Spermidine/Spermine N-Acetyltransferases in *Staphylococcus Aureus* and *Enterococcus Faecalis*. *Mol. Microbiol.* **2019**, *111*, 159–175. [[CrossRef](#)]
79. Campbell, A.G.; Schwientek, P.; Vishnivetskaya, T.; Woyke, T.; Levy, S.; Beall, C.J.; Griffen, A.; Leys, E.; Podar, M. The Under-Recognized Dominance of Verrucomicrobia in Soil Bacterial Communities. *Environ. Microbiol.* **2015**, *16*, 2635–2643. [[CrossRef](#)]
80. Tláškal, V.; Voříšková, J.; Baldrian, P. Bacterial Succession on Decomposing Leaf Litter Exhibits a Specific Occurrence Pattern of Cellulolytic Taxa and Potential Decomposers of Fungal Mycelia. *FEMS Microbiol. Ecol.* **2016**, *92*, fiw177. [[CrossRef](#)] [[PubMed](#)]
81. van der Lelie, D.; Taghavi, S.; McCorkle, S.M.; Li, L.-L.; Malfatti, S.A.; Monteleone, D.; Donohoe, B.S.; Ding, S.-Y.; Adney, W.S.; Himmel, M.E.; et al. The Metagenome of an Anaerobic Microbial Community Decomposing Poplar Wood Chips. *PLoS One* **2012**, *7*, e36740. [[CrossRef](#)] [[PubMed](#)]
82. Mazzoli, R. Development of Microorganisms for Cellulose-Biofuel Consolidated Bioprocessings: Metabolic Engineers' Tricks. *Comput. Struct. Biotechnol. J.* **2012**, *3*, e201210007. [[CrossRef](#)] [[PubMed](#)]
83. Hutchinson, H.B.; Clayton, J. On the Decomposition of Cellulose by an Aerobic Organism (*Spirochaeta cytophaga*, n. sp.). *J. Agric. Sci.* **1919**, *9*, 143–172. [[CrossRef](#)]
84. Gladkov, G.; Kimeklis, A.; Zverev, A.; Pershina, E.; Ivanova, E.; Kichko, A.; Andronov, E.; Abakumov, E. Soil Microbiome of the Postmining Areas in Polar Ecosystems in Surroundings of Nadym, Western Siberia, Russia. *Open Agric.* **2019**, *4*, 684–696. [[CrossRef](#)]
85. Callahan, B.J.; McMurdie, P.J.; Rosen, M.J.; Han, A.W.; Johnson, A.J.A.; Holmes, S.P. DADA2: High-Resolution Sample Inference from Illumina Amplicon Data. *Nat. Methods* **2016**, *13*, 581–583. [[CrossRef](#)]
86. R Core Team. R: A Language and Environment for Statistical Computing. Available online: <https://www.R-project.org/> (accessed on 12 August 2022).
87. Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glöckner, F.O. The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools. *Nucleic Acids Res.* **2013**, *41*, D590–D596. [[CrossRef](#)]
88. Nilsson, R.H.; Larsson, K.-H.; Taylor, A.F.S.; Bengtsson-Palme, J.; Jeppesen, T.S.; Schigel, D.; Kennedy, P.; Picard, K.; Glöckner, F.O.; Tedersoo, L.; et al. The UNITE Database for Molecular Identification of Fungi: Handling Dark Taxa and Parallel Taxonomic Classifications. *Nucleic Acids Res.* **2019**, *47*, D259–D264. [[CrossRef](#)]
89. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; von Haeseler, A.; Lanfear, R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **2020**, *37*, 1530–1534. [[CrossRef](#)]
90. McMurdie, P.J.; Holmes, S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* **2013**, *8*, e61217. [[CrossRef](#)]
91. Andersen, K.S.; Kirkegaard, R.H.; Karst, S.M.; Albertsen, M. Ampvis2: An R Package to Analyse and Visualise 16S RRNA Amplicon Data. *bioRxiv* **2018**. [[CrossRef](#)]
92. Shannon, C.E.; Weaver, W. *The Mathematical Theory of Communication*; University of Illinois Press: Champaign, IL, USA, 1949.
93. Simpson, E.H. Measurement of Diversity. *Nature* **1949**, *163*, 688. [[CrossRef](#)]
94. Mann, H.B.; Whitney, D.R. On a Test of Whether One of Two Random Variables Is Stochastically Larger than the Other. *Ann. Math. Stat.* **1947**, *18*, 50–60. [[CrossRef](#)]
95. Gower, J.C. Principal Coordinates Analysis. In *Wiley StatsRef: Statistics Reference Online*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2015; pp. 1–7.
96. Bray, J.R.; Curtis, J.T. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecol. Monogr.* **1957**, *27*, 325–349. [[CrossRef](#)]

97. Silverman, J.D.; Washburne, A.D.; Mukherjee, S.; David, L.A. A Phylogenetic Transform Enhances Analysis of Compositional Microbiota Data. *eLife* **2017**, *6*, e21887. [[CrossRef](#)]
98. Wick, R.R.; Judd, L.M.; Holt, K.E. Performance of Neural Network Basecalling Tools for Oxford Nanopore Sequencing. *Genome Biol.* **2019**, *20*, 129. [[CrossRef](#)]
99. Kolmogorov, M.; Bickhart, D.M.; Behsaz, B.; Gurevich, A.; Rayko, M.; Shin, S.B.; Kuhn, K.; Yuan, J.; Polevikov, E.; Smith, T.P.L.; et al. MetaFlye: Scalable Long-Read Metagenome Assembly Using Repeat Graphs. *Nat. Methods* **2020**, *17*, 1103–1110. [[CrossRef](#)]
100. Vaser, R.; Sović, I.; Nagarajan, N.; Šikić, M. Fast and Accurate de Novo Genome Assembly from Long Uncorrected Reads. *Genome Res.* **2017**, *27*, 737–746. [[CrossRef](#)]
101. Medaka: Sequence Correction Provided by ONT Research. Available online: <https://github.com/nanoporetech/medaka> (accessed on 22 February 2022).
102. Wood, D.E.; Lu, J.; Langmead, B. Improved Metagenomic Analysis with Kraken 2. *Genome Biol.* **2019**, *20*, 257. [[CrossRef](#)]
103. Struo2: A Pipeline for Building Custom Databases for Common Metagenome Profilers. Available online: <https://github.com/leylabmpi/Struo2> (accessed on 12 August 2022).
104. Loman Lab Mock Community Experiments. Databases. Available online: https://lomanlab.github.io/mockcommunity/mc_databases.html (accessed on 12 August 2022).
105. Ondov, B.D.; Bergman, N.H.; Phillippy, A.M. Interactive Metagenomic Visualization in a Web Browser. *BMC Bioinform.* **2011**, *12*, 385. [[CrossRef](#)]
106. Yin, Y.; Mao, X.; Yang, J.; Chen, X.; Mao, F.; Xu, Y. DbCAN: A Web Resource for Automated Carbohydrate-Active Enzyme Annotation. *Nucleic Acids Res.* **2012**, *40*, W445–W451. [[CrossRef](#)] [[PubMed](#)]
107. Buchfink, B.; Reuter, K.; Drost, H.-G. Sensitive Protein Alignments at Tree-of-Life Scale Using DIAMOND. *Nat. Methods* **2021**, *18*, 366–368. [[CrossRef](#)] [[PubMed](#)]
108. Busk, P.K.; Pilgaard, B.; Lezyk, M.J.; Meyer, A.S.; Lange, L. Homology to Peptide Pattern for Annotation of Carbohydrate-Active Enzymes and Prediction of Function. *BMC Bioinform.* **2017**, *18*, 214. [[CrossRef](#)]
109. Huang, L.; Zhang, H.; Wu, P.; Entwistle, S.; Li, X.; Yohe, T.; Yi, H.; Yang, Z.; Yin, Y. DbCAN-Seq: A Database of Carbohydrate-Active Enzyme (CAZyme) Sequence and Annotation. *Nucleic Acids Res.* **2018**, *46*, D516–D521. [[CrossRef](#)] [[PubMed](#)]
110. Cantalapiedra, C.P.; Hernández-Plaza, A.; Letunic, I.; Bork, P.; Huerta-Cepas, J. EggNOG-Mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol. Biol. Evol.* **2021**, *38*, 5825–5829. [[CrossRef](#)]
111. Huerta-Cepas, J.; Szklarczyk, D.; Heller, D.; Hernández-Plaza, A.; Forslund, S.K.; Cook, H.; Mende, D.R.; Letunic, I.; Rattei, T.; Jensen, L.J.; et al. EggNOG 5.0: A Hierarchical, Functionally and Phylogenetically Annotated Orthology Resource Based on 5090 Organisms and 2502 Viruses. *Nucleic Acids Res.* **2019**, *47*, D309–D314. [[CrossRef](#)] [[PubMed](#)]
112. Wu, T.; Hu, E.; Xu, S.; Chen, M.; Guo, P.; Dai, Z.; Feng, T.; Zhou, L.; Tang, W.; Zhan, L.; et al. ClusterProfiler 4.0: A Universal Enrichment Tool for Interpreting Omics Data. *Innovation* **2021**, *2*, 100141. [[CrossRef](#)] [[PubMed](#)]
113. Fisher, R.A. *Statistical Methods for Research Workers*; Oliver and Boyd: Edinburgh, UK, 1938.
114. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **1995**, *57*, 289–300. [[CrossRef](#)]
115. Oksanen, J.; Blanchet, F.G.; Kindt, R.; Legendre, P.; Minchin, P.R.; O’Hara, R.B.; Simpson, G.L.; Solymos, P.; Stevens, M.H.H.; Wagner, H. Vegan: Community Ecology Package. R Package Version 2.2-0. 2014. Available online: <http://CRAN.Rproject.org/package=vegan> (accessed on 12 August 2022).
116. Kang, D.D.; Li, F.; Kirton, E.; Thomas, A.; Egan, R.; An, H.; Wang, Z. MetaBAT 2: An Adaptive Binning Algorithm for Robust and Efficient Genome Reconstruction from Metagenome Assemblies. *PeerJ* **2019**, *7*, e7359. [[CrossRef](#)]
117. Parks, D.H.; Imelfort, M.; Skennerton, C.T.; Hugenholtz, P.; Tyson, G.W. CheckM: Assessing the Quality of Microbial Genomes Recovered from Isolates, Single Cells, and Metagenomes. *Genome Res.* **2015**, *25*, 1043–1055. [[CrossRef](#)]
118. Parks, D.H.; Chuvochina, M.; Rinke, C.; Mussig, A.J.; Chaumeil, P.-A.; Hugenholtz, P. GTDB: An Ongoing Census of Bacterial and Archaeal Diversity through a Phylogenetically Consistent, Rank Normalized and Complete Genome-Based Taxonomy. *Nucleic Acids Res.* **2022**, *50*, D785–D794. [[CrossRef](#)]
119. Shaffer, M.; Borton, M.A.; McGivern, B.B.; Zayed, A.A.; La Rosa, S.L.; Solden, L.M.; Liu, P.; Narrowe, A.B.; Rodríguez-Ramos, J.; Bolduc, B.; et al. DRAM for Distilling Microbial Metabolism to Automate the Curation of Microbiome Function. *Nucleic Acids Res.* **2020**, *48*, 8883–8900. [[CrossRef](#)] [[PubMed](#)]