Metagenomics Study To Compare The Taxonomy And Metabolism of a Lignocellulolytic Microbial Consortium Cultured in Different Carbon Conditions

Qinggeer BORJIGIN (qinggeer1202@sina.com)
Inner Mongolia Agricultural University https://orcid.org/0000-0002-7071-1006

Bizhou ZHANG
Inner Mongolia Academy of Agricultural and Animal Husbandry Sciences

Xiaofang Yu
Inner Mongolia Agriculture University: Inner Mongolia Agricultural University

Julin Gao
Inner Mongolia Agricultural University

Xin ZHANG
Inner Mongolia Agricultural University

Jiawei Qu
Inner Mongolia Agricultural University

Daling Ma
Inner Mongolia Agricultural University

Shuping Hu
Inner Mongolia Agricultural University

Shengcai Han
Inner Mongolia Agricultural University

Research Article

Keywords: Functional potentials, Metagenomics, Microbial consortium, Taxonomic composition

Posted Date: December 1st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1076547/v1

Abstract

A lignocellulolytic microbial consortium holds promise for the in situ biodegradation of crop straw and the comprehensive and effective utilization of agricultural waste. In this study, we applied metagenomics technology to comprehensively explore the metabolic functional potential and taxonomic diversity of the microbial consortia CS (cultured on corn stover) and FP (cultured on filter paper).Analyses of the metagenomics taxonomic affiliation data showed considerable differences in the taxonomic composition and functional profile of the microbial consortia CS and FP. The microbial consortia CS primarily contained members from the genera *Pseudomonas, Stenotrophomonas, Achromobacter, Dysgonomonas*, *Flavobacterium* and *Sphingobacterium*, as well as *Cellvibrio, Azospirillum, Pseudomonas, Dysgonomonas* and *Cellulomonas* in FP. The COG and KEGG annotation analyses revealed considerable levels of diversity. Further analysis determined that the CS consortium had an increase in the acid and ester metabolism pathways, while carbohydrate metabolism was enriched in the FP consortium. Furthermore, a comparison against the CAZy database showed that the microbial consortia CS and FP contain a rich diversity of lignocellulose degrading families, in which GH5, GH6, GH9, GH10, GH11, GH26, GH42, and GH43 were enriched in the FP consortium, and GH44, GH28, GH2, and GH29 increased in the CS consortium. The degradative mechanism of lignocellulose metabolism by the two microbial consortia is similar, but the annotation of quantity of genes indicated that they are diverse and vary greatly. The lignocellulolytic microbial consortia cultured under different carbon conditions (CS and FP) differed substantially in their composition of the microbial community at the genus level. The changes in functional diversity were accompanied with variation in the composition of microorganisms, many of which are related to the degradation of lignocellulolytic materials. The genera Pseudomonas, Dysgonomonas and Sphingobacterium in CS and the genera Cellvibrio and Pseudomonas in FP exhibited a much wider distribution of lignocellulose degradative ability.

Introduction

As renewable resources of energy, plant biomass has a substantial potential to produce alternative liquid fuels, and lignocellulose is the main component of the plant cell wall (Limayem et al.2012). Lignocellulose is composed of cellulose and hemicellulose polysaccharides and the aromatic polymer lignin (López-Mondéjar et al.2016). However, inefficient lignocellulose deconstruction is a prime bottleneck for its economic conversion and further utilization (Chang et al. 2000). In addition to the chemical and/or physical pretreatment that breaks down the complex bonds that link lignocellulosic materials, the action of effective multiple biomass-degrading enzymes is also required (Lemos et al. 2017). Natural or engineered microbial consortia that have many different enzymes that act synergistically are a promising way to overcome the heterogeneity and complexity of lignocellulose (Sheng et al.2016; Minty et al. 2013; Cortes-Tolalpa et al. 2016; Chandel et al.2011; Wang et al. 2016; Lewin et al. 2016).

The taxonomic composition and functional profile of the microbial consortia that may even act synergistically and be interdependent with each other are primarily influenced by the cultural conditions, and cultivation and screening in material rich in lignocellulose is a standard approach to screen efficient lignocellulose-adapted microbial consortia (Allgaier et al.2010; de Lima Brossi et al.2016). For example, Jiménez et al.(Jiménez et al.2016) compared microbial consortia selected from the same source of forest soil and subsequently subcultured them on wheat straw, xylan, and xylose, which revealed that the key microbes in the three final

consortia differed. Jiménez and Maruthamuthu (Maruthamuthu et al.2016; Jiménez et al.2015) demonstrated a discrepancy in the metabolism and functional cloned genes of two lignocellulolytic microbial consortia RWS (cultured on untreated wheat straw) and TWS (cultured on heat-treated wheat straw) derived from forest soil. Carlos et al. (Carlos et al.2018) found that the ability of microbial consortia cultured on xylan and alkali lignin media to degrade lignocellulose exhibited extreme differences in their community microbial composition and enzymes after sub-culturing, independently of the original consortium.

In a previous study, we obtained a microbial consortium from corn straw returned soils by restrictive cultivation and cold acclimation (Qinggeer et al.2016). The GF-20 consortium remained relatively stable with regard to Lignocellulose decomposition, growth, and population structure, even with an increase in the number of subcultures. Metagenomics is a promising approach to explore the functional potential of microbiomes. We used this approach to analyze the composition and functional changes and the lignocellulolytic mechanism present in the GF-20 consortium, which revealed changes in the abundance of distinct microbial members in response to incubation on corn straw and filter paper. In addition, we identified metabolic pathways that were enriched after the substrate was changed. The data from this study improve our comprehension of the ecology of degradation of straw and straw-based selection. They also provide insight into the metabolic potential present in the resulting microbial consortia, opening up possibilities for further ecological and biotechnological studies.

Materials And Methods

Lignocellulolytic microbial consortia

The microbial consortia GF-20 had been reported before (Yu et al.2019). Briefly, cell suspensions were prepared by adding 10 g of sampled soil to 250 ml flasks that contained 100 ml of AOM media. The flasks were incubated for 2 d in a shaker incubator at 100 rpm and then incubated for 13 d in a static condition. When the filter paper had been degraded to pieces, 10% (v/v) of well-blended enriched fluid was transferred to new AOM media. The temperature was decreased by 1°C until each successive transfer of 1-2 generations reached 10°C in a static condition. Finally, the microbial consortium GF-20, which efficiently degraded lignocellulose and possessed a stable microbial composition, was obtained by analyzing the activities of the degradation of corn stover and the activities of cellulase. The cell extracts (1 ml) were then inoculated in triplicate flasks that contained 20 ml of AOM media with 2 g of raw corn stover (CS) or filter paper (FP). The flasks were cultured at 10°C with static conditions. Subsequently, 1 ml aliquots were transferred to 20 ml of fresh AOM when the systems reached 7-8 log cells/ml. The triplicate CS and FP cultures were sampled after 10 transfers. The selection of samples was based on previous bacterial PCR-DGGE results (Yu et al.2019).

pH and cellulase activity measurement

Microbial consortium CS and FP obtained by subculturing under different carbon conditions was inoculated on the AOM medium at 10°C with Static conditions for 12 d. The cultures were sampled every two days for determination of pH and cellulase activity. A 0.5-mL sample was taken from the solution for pH measurement. Fermentation samples (5 mL) were centrifuged at 12000 rpm at 4°C for 10 min, and the supernatants were used for determination of cellulase activity. The pH was measured with a pH meter (HORIBA B-712, Japan), whereas the cellulase activity was determined by the dinitrosalicylic acid (DNS) method at pH 4.8 and 50°C

(using the method recommended by IUPAC) (Ghose et al. 1987). One unit (U) of enzymatic activity was defined as the amount of enzyme that could convert 1 μmol cellulosic substrate within 1 minute.

Metagenome sequencing

The genomic DNA was extracted from the CS and FP samples (n=6) using an E.Z.N.A.® soil DNA Kit (Omega Bio-tek, Norcross, GA, USA) according to the manufacturer's instructions. The DNA extract was monitored using 1.0% agarose gel electrophoresis, and a NanoDrop 2000 UV-Vis spectrophotometer (ThermoFisher Scientific, Waltham, MA, USA) was used to determine the concentration and purity of DNA. The DNA was fragmented to approximately 400 bp by Covaris M220 for a constructed PE library using NEXTFLEX Rapid DNA-Seq (Bioo Scientific, Austin, TX, USA). PE sequencing was performed on an Illumina HiSeq platform (Illumina, Inc., San Diego, CA, USA) at Majorbio Bio-Pharm Technology Co., Ltd. (Shanghai, China) using HiSeq X Reagent Kits. The raw reads were deposited into the NCBI Sequence Read Archive (SRA) database (Accession Number: SRA 272980).

Sequence quality control and genome assembly

Adapter sequences were stripped from the 3' and 5' ends of paired-end Illumina reads using SeqPrep (https://github.com/jstjohn/SeqPrep). Low quality reads with a quality value < 20 or length < 50 bp or containing N bases were removed by Sickle (https://github.com/najoshi/sickle). Contigs ≥ 300 bp were selected and used for further gene prediction and annotation.

Gene prediction, taxonomy, and functional annotation

The predicted open reading frames (ORFs) with a length ≥ 100 bp were selected and translated into amino acid sequences. Use MetaGene (Noguchi et al.2006) (http://metagene.cb.k.u-tokyo.ac.jp/) to predict the ORF of contigs in the splicing result. The predicted gene sequences of all the samples were clustered, and the longest gene in each category was considered to be the representative sequence. The CD-HIT tool (http://www.bioinformatics.org/cd-hit/) was used to construct a non-redundant gene set(95% identity,90% coverage). In addition, the gene abundance information of sample was counted using SOAPaligner software (http://soap.genomics.org.cn/).

The taxonomic annotations used BLASTp (Altschul et al.1997) to compare the non-redundant gene set with the NR database with an E-value cutoff of 10^{-5} and calculate the relative abundance (RA) of the species in the Domain, Kingdom, Phylum, Class, Order, Family, Genus, and Species. The abundance of species in each sample was counted at each taxonomic level. Cluster of Orthologous Groups ofproteins (COG) and (Kyoto Encyclopedia of Genes and Genomes) KEGG annotation for the representative sequences was performed using BLASTp (v. 2.2.28+) against the Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups (eggnog) database (Jensen et al.2008; Tatusov et al.2003) (v. 4.5) and the KEGG database (Kanehisa et al.2000). According to the comparison result, use KOBAS 2.0 (KEGG Orthology Based Annotation System, http://kobas.cbi.pku.edu.cn/home.do) for functional annotation (Xie et al.2011). Use the sum of gene abundance corresponding to KO, Pathway, and Module to calculate the abundance of the corresponding functional category. Carbohydrate-active enzymes annotation was conducted using hmmscan (http://hmmer.janelia.org/search/hmmscan) against the CAZy database v. 5.0 (Lombard et al.2014). All the annotation had an E-value cutoff of $1e^{-5}$.

Statistical analysis

To evaluate the relative abundance (RA) of reads per phylum, genus and CAZy family, the counts were normalized using the total numbers of quality reads (matched in the database) per metagenome. Moreover, the RA of reads of each phylum, genus and CAZy family was compared between the metagenomes from the selected consortia. Fold changes were calculated based on the normalized RA values (fold increase= CS consortia/FP) and then Log10 transformed Difference analysis was all conducted with the “stats” package in R (v. 3.6.0). STAMT was used to conduct a pair-wise statistical comparison of the KEGG and COG classification (Parks et al. 2010).Two-sided Fisher’s and/or ANOVA tests were used for hypothesis testing. Differences in proportions and 95% confdence intervals (CIs) were calculated according to the Newcombe-Wilson method. Multiple test corrections were done using the Storey FDR q-value false discovery rate. SigmaPlot (v.12.5) used to visualize the abundance of phylum-level species and CAZymes at the class level. According to the abundance of KEGG prthology, use LEfSe (http://huttenhower.sph.harvard.edu/galaxy/root?tool_id=lefse_upload) to perform linear discriminant analysis (LDA) according to different grouping conditions. The BLAST results were loaded into MEGAN (v5) sofware and classifed taxonomically according to suggested parameters for the Lowest Common Ancestor algorithm (LCA) (maximum number of matches per read: 10; min support: 5; min score: 35; max expected 0.01; min complexity 0.3; and top percent: 10 (Huson et al. 2011). The co-occurrence network of genus (detected other microbial communities) was visualized by Cytoscape (v. 3.6.0). To evaluate the relative abundance (RA) of reads CAZy per family, the counts were normalized using the total numbers of quality reads (matched in the database) per metagenome. Moreover, the RA of reads of each class (AA, CBM, CE, GH, GT and PL) and family was compared between the metagenomes from the selected consortia (1W, 3W, 10W, 1T, 3T and 10T) as well as the FS1. Fold changes were calculated based on the normalized RA values (fold increase= RA consortia/CS of FP) and then Log10 transformed.

Results

Metagenomic reference

Metagenomic sequencing of the microbial consortia CS and FP generated a total of 47,206,833 and 48,912,993 reads (N50 length of 23,162 and 17,589 bp, respectively), which were then assembled into 51,160 and 80,221 contigs, respectively. The assembly rates were 40.67% and 66.43%, respectively (Table 1).

Table1 Summary of Illumina GA reads and ORF prediction

Sample ID	Raw Reads	Clean reads	Clean reads rate (%)	Contigs	Contigs bases	N50	Assembly results (%)	Number of genes
CS	47,206,833	46,442,751	98.39	51,160	11,2661,346	23,162	40.67	127,794
FP	48,912,993	48,405,294	98.96	80,221	97,884,307	17,589	66.43	183,467

Taxonomic composition in the microbial consortia

The analysis of metagenomics taxonomic annotation showed that GF-20 was predominantly populated with bacteria (99.93% in CS and 99.11% in FP of total sequences), along with a small number of eukaryotes and archaea. The assignable sequences indicated that 78.13% and 80.62% of the microbes were affiliated with Proteobacteria in CS and FP, respectively, followed by Bacteroidetes, Actinobacteria and Firmicutes. The relative abundance of Bacteroidetes in CS was significantly higher than that of FP (1.98-fold RA increase), and the relative abundances of Actinobacteria and Firmicutes in FP were higher than those in CS (8.35, 5.8-fold RA increase, respectively) (Fig. 1a).

Taxonomic annotation was performed at the genus level to gain further insight into the microbial composition diversity within GF-20 that was cultured under corn stover and filter paper carbon conditions. *Pseudomonas* (60.59%), *Stenotrophomonas* (4.54%), *Achromobacter* (5.08%), *Dysgonomonas* (7.60%), *Flavobacterium* (3.51%), and *Sphingobacterium* (6.02%) were extremely abundant in the CS consortia, while the *Achromobacter*, *Dysgonomonas*, and *Cellulomonas* genera were present at low abundance in the FP consortia. A total of 23.80%, 12.86%, 12.66%, 4.65%, 1.35%, 1.33%, 5.16%, and 5.28% of the predicted microbes were affiliated with *Cellvibrio*, *Azospirillum*, *Pseudomonas*, *Acidovorax*, *Hydrogenophaga*, *Stenotrophomonas*, *Dysgonomonas*, and *Cellulomonas* in FP, respectively. However, the relative abundances of those genera in FP were significantly higher than those in CS, with the exception of *Pseudomonas*, *Stenotrophomonas*, and *Dysgonomonas*(Fig. 1b). Based on the distribution of microorganisms in the microbial consortia CS and FP, the microbial composition diversity of GF-20 had diversified to an extreme extent when cultured under different carbon substrates.

Metabolic functional profile in the microbial consortia

To assess the metabolic functional diversity of the microbial consortia CS and FP, the COG and KEGG databases were used for annotation. BLAST comparisons generated 200,261 hits for the COG database and 133,053 hits for the KEGG one. Based on comparative and pairwise analyses of the COG-based functional profiles, CS and FP correlated at r^2=0.638 (Fig. 2a). The COG ontology analysis is shown in Fig. 2b, in which Category of metabolism (39.70±0.26% in CS and 36.29±0.16% in FP) was the dominant function among the 23 categories. Notably, metabolism was related to the growth of microbial community. Remarkably, the relative abundance of Inorganic ion transport and metabolism, Amino acid transport and metabolism, Carbohydrate transport and metabolism, Energy production and conversion, Lipid transport and metabolism and Nucleotide transport and metabolism in CS was significantly higher than that in FP (1.23-, 1.29-, 1.09-, 1.09-, 1.15-, and 1.04-fold RA increase, respectively). The matrix of heteropolymeric substrates determines the complexity of its metabolic functions, thus, enriching the metabolic functions related to the decomposition of lignocellulose in CS.

An analysis of the predicted metabolic profile of the metagenomics data set was conducted for functions using the KEGG databases. The functional annotation of genes revealed that no significant difference in pathways, modules, orthologies, and enzymes existed between CS and FP (Table 2). The expressed genes were used to predict the changes in metabolic function (Fig. 3 and Fig. 4). Comparative KEGG-based analyses of the microbial consortia CS showed that the pathways, modules, and KEGG Orthologies still revealed considerable levels of similarity with those of FP (r^2=0.955, 0.708, 0.774, Fig. 3). Thus, both consortia exhibited a greatly similar functional profile. KEGG-based analyses of differentially annotated genes between CS and FP are shown in Fig. 4, which illustrates that 11.62% and 15.92% of the genes were involved in carbohydrate

metabolism, respectively, 12.58% and 12.55% in amino acid metabolism, respectively, and 10.11% and 12.18% in global and overview maps, respectively. These metabolic pathways were the most enriched.

Table2 Summary of the KEGG annotation results

Annotation categories	KEGG Genes	KEGG Pathway	KEGG Module	KEGG Orthology	Enzymes
CS	28,113,438a	396a	498a	5,898a	1,915a
FP	21,347,752a	399a	490a	5,931a	1,960a

Note: Different letters indicate significant differences at the 0.05 level.

We used the Lefse tool to analyze the genes and pathways that consistently explain the differences between CS and FP. A total of 34 KEGG pathways (Fig. 5) and 95 KEGG orthology groups (KO) (Supplementary Fig. 1) differed significantly between CS and FP. Microbial communities from the CS consortia had an increase in acid and ester metabolism pathways, such as Tryptophan metabolism (ko00380), Glyoxylate and dicarboxylate metabolism (ko00630),Valine, leucine and isoleucine degradation (ko00280), and Butanoate metabolism (ko00650) among others. In contrast, the FP consortia were enriched in carbohydrate metabolism, such as Starch and sucrose metabolism (ko00500), and Fructose and mannose metabolism (ko00051) among others. Interestingly, a two-component system (ko02020) that has a domain in the periplasm to sense carbohydrates and regulate (hemi) cellulose utilization loci (HULs) was present at a significantly higher abundance in CS (Fig. 5). 3 (ko) between CS (green) and FP (red) individuals (LDA>3).

Diversity profile of CAZymes in microbial consortia

To detect the genes that contribute to the degradation of lignocellulose, we performed a BLASTx of all the metagenome reads against the CAZy database and identified the number of different CAZymes that were present in the microbial consortia and analyzed the influence of different types of carbon sources on the proportion of enzymes (Fig. 6). The CS and FP consortia contained 5,974±667 and 8,420±415 different CAZyme genes, respectively, that were distributed unequally between glycoside hydrolases (GHs, 45.58±0.93% and 41.51±0.89%, respectively), carbohydrate-binding modules (CBMs, 5.73±0.14% and 13.66±0.39%, respectively), polysaccharide lyases (PLs, 1.58±0.11% and 3.68±0.18%, respectively), carbohydrate esterases (CEs, 16.81±0.45% and 12.34±0.02%, respectively), glycosyltransferases (GTs, 23.37±0.34% and 24.22±0.89%, respectively), and auxiliary activities (AAs, 6.92±0.39% and 4.60±0.10%, respectively). Remarkably, the relative abundances of GHs, CEs and AAs in CS were significantly higher than those in FP (1.10-, 1.36-, 1.50-fold RA increase, respectively), while the CBMs and PLs in FP were significantly higher than those in CS (2.38- and 2.32-fold RA increase, respectively). It is notable that the GHs and GTs genes comprise a relatively high proportion (68.96% and 65.72%, respectively).

There is some difference in the rate of contribution of the microbial consortia CS and FP to CAZyme at the phylum and genus levels (Supplementary Fig. 2). The difference in analysis of reads of CAZy-GH families in the CS and FP consortia found that the reads of GH2, GH109, GH92, GH29, GH88, GH114, GH110, GH32, GH125, GH116, GH65, GH56, GH4, and GH89 in CS showed more richness than those in FP. However, the reads of

GH18, GH9, GH15, GH81, GH74, GH105, GH11, GH115, GH43, GH35, GH27, GH57, and GH5 were enriched in FP (Fig. 7a). Based on the taxonomic classification by the lowest common ancestor (LCA) algorithm (Fig. 7b and c), the GH2, GH92, GH29, GH125, and GH116 reads of CS were primarily assigned to *Dysgonomonas* and *Sphingobacterium*, and the ranges of contribution ranges were 25.87-73.50% and 17.84-57.28%, respectively. The GH109, GH88, and GH32 families were primarily affiliated with organisms from within the *Pseudomonas*, *Dysgonomonas* and *Sphingobacterium* genera; the GH114 families were primarily affiliated with *Pseudomonas* (96.58%), and GH89 was primarily affiliated with *Dysgonomonas* (96.94%). The reads of glycoside hydrolases enriched in FP differed significantly from those of CS in the contribution of microbial organisms. Figure 7c indicates that the reads assigned to all the enriched GH families (with the exception of GH35, GH27, and GH5) in FP were primarily affiliated with *Cellvibrio*, and its rate of contribution was between 30%-85%. In our study, the phyla Proteobacteria, Bacteroidetes, and Actinobacteria harbored more CAZyme genes, and the genera *Pseudomonas* and *Dysgonomonas* in CS and the genera *Cellvibrio* and *Pseudomonas* in FP exhibited a much wider distribution of the CAZyme repertoire.

The families of glycosylic hydrolase enzymes that were detected represent known carbohydrate-active enzymes, including cellulases, hemicellulases, oligosaccharide degrading enzymes, and debranching enzymes, that are pivotal to the degradation of carbohydrate molecules. A total of 2,067 and 2,725 GH-related sequences in CS and FP were identified in the metagenome accounting, in which a sum of 29 different families related to lignocellulose degradation was found. These families represent 31.18% and 39.43% of the GH genes identified in each community, respectively. Each of the GH families detected in the consortia CS metagenome was also present in the consortia FP, but there were significant differences in their relative abundances. The abundances of cellulases (GH5, GH6, GH7, GH9, GH44, GH45, and GH48) in the CS and FP consortia were 2.56% and 9.25%, respectively. The content of hemicellulases (GH8, GH10, GH11, GH12, GH26, GH28, and GH53) comprised 3.03% and 5.86%, respectively. The debranching enzymes (GH51, GH54, GH62, GH67, and GH78) comprised 2.71% and 3.29%, respectively, and oligosaccharide degrading enzymes (GH1, GH2, GH3, GH29, GH35, GH38, GH39, GH42, GH43, and GH52) comprised 22.88 and 21.03% in CS and FP, respectively (Table 3). It is worth noting the enrichments of GH5 (cellulases), GH6, (endoglucanases), GH9 (endoglucanases), GH10 (endo-1,4-β-xylanases), GH11 (xylanases), GH26 (β-mannanase & xylanases), GH42 (β-galactosidases), GH43 (arabino/xylosidases) in the FP consortia and GH44 (endoglucanases), GH28 (galacturonases), GH2 (β-galactosidases), GH29 (α-L-fucosidases) in the CS consortia. During the culture period, the enzyme activities of FPA, Cx, CBH and CB showed that CS was significantly higher than FP (Fig. 8). The reads of lignocellulose degradation related-GH families indicated that GH3, GH2, GH8, GH28, GH38, GH78, and GH29 in CS were significantly higher than those in FP. The GH2, GH3, GH8, GH78, and GH38 families were mostly affiliated with organisms from within the *Dysgonomonas* genus of Bacteroides (52.16%, 61.53%, 87.87%, 50.75%, and 67.47%, respectively). The GH29 families were primarily found in the *Dysgonomonas* and *Sphingobacterium* genera of Bacteroides (44.22% and 21.98%, respectively). The GH28 families were primarily associated with *Pedobacter* (59.67%)(Fig. 9).

Table 3
Comparative relative abundances of genes that encode lignocellulose degradative enzyme homologs in different environmental metagenomes with a high potential to degrade lignocellulose

	Predominant activity	CS	FP	RSA[a]	BGC-1[b]	SAC[c]	reactor[d]	Compost[e]
Cellulases								
GH5	Cellulases	0.92	5.56	2.36	2.28	3.2	3.12	2.58
GH6	Endoglucanases	0.04	0.55	0.77	/	2.1	0	0.94
GH7	Endoglucanases	/	/	0	/	0.1	0	0.17
GH9	Endoglucanases	0.61	2.61	1.12	2.59	4.3	1.22	2.92
GH44	Endoglucanases	0.88	0.34	0.06	0.31	0.4	0.08	0
GH45	Endoglucanases	0.07	0.00	0	0.00	0	0	0.09
GH48	Cellobiohydrolases	0.03	0.19	0.3	/	0.5	1.07	0.09
Subtotal		2.56	9.25	4.61	5.19	10.60	5.49	6.79
Hemicellulases								
GH8*	Endoxylanases	0.13	0.08	0	0.41	0.5	0.15	0.69
GH10	Endo-1,4-β-xylanases	0.80	2.07	2.18	2.28	8.9	3.65	4.21
GH11	Xylanases	0.03	1.00	0.47	0.41	1.4	0.46	0.6
GH12	Xyloglucanases	0.01	0.05	0.41	0.72	0.6	0	0.34
GH26	β-mannanase & xylanases	0.56	1.33	0.83	/	1.5	1.6	1.63
GH28*	Galacturonases	1.49	0.72	0.65	2.28	0.9	1.07	0.77
GH53	Endo-1,4-β-galactanases	0.03	0.62	0.35	0.31	0.2	1.29	0.26
Subtotal		3.03	5.86	4.89	6.41	14.00	8.22	8.50
Oligosaccharide degrading enzymes								
GH1	β-glucosidases	0.13	1.08	3.72	0.73	9.2	8.9	6.78
GH2*	β-galactosidases	6.54	3.36	1.59	9.60	8.6	4.41	3.78
GH3*	β-glucosidases	8.28	7.42	4.9	7.41	12.2	7.31	7.04
GH29*	α-L-fucosidases	2.81	0.47	0.89	1.67	2.1	1.9	2.23
GH35	β-galactosidases	0.19	0.62	0.41	0.31	0.6	0.84	0.52

*Indicates that CS is significantly higher than FP at the 0.05 level. *RSA*[a] : Wang et al., 2016. *BGC-1*[b]: *Wongwilaiwalin et al.,2013. SAC*[c]: Allgaier et al., 2010. *reactor*[d]: Hollister et al., 2012. *Compost*[e]: Allgaier et al., 2010.

	Predominant activity	CS	FP	RSA[a]	BGC-1[b]	SAC[c]	reactor[d]	Compost[e]
GH38*	α-mannosidases	0.39	0.19	0.77	0.73	2.6	1.98	1.55
GH39	β-xylosidases	0.17	0.25	1.89	1.15	1	2.13	0.94
GH42	β-galactosidases	0.03	0.22	0.83	0.42	2.5	1.98	1.89
GH43	Arabino/xylosidases	4.33	7.42	3.01	9.08	11.3	5.78	7.81
GH52	β-xylosidases	0.00	0.00	0.18	0.31	0	0.53	0
Subtotal		22.88	21.03	18.19	31.40	50.10	35.76	32.54
Debranching enzymes								
GH51	α-L-arabinofuranosidases	0.74	1.05	1.42	2.21	7.8	3.96	1.03
GH54	α-L-arabinofuranosidases	/	/	0	0.00	0	0	0
GH62	α-L-arabinofuranosidases	0.03	0.45	0.35	0.11	1.7	0	0.43
GH67	α-glucuronidases	0.36	0.73	0.35	0.84	3.6	0.91	2.06
GH78*	α-L-rhamnosidases	1.58	1.06	2.3	3.05	8.1	1.06	4.72
Subtotal		2.71	3.29	4.42	6.20	21.20	5.93	8.24
Total		31.18	39.43	32.11	49.20	95.90	55.40	56.07
Total GHs		2067	2725	1694	957	801	1314	27755

*Indicates that CS is significantly higher than FP at the 0.05 level. *RSA*[a] : Wang et al., 2016. *BGC-1*[b]: *Wongwilaiwalin et al.,2013. SAC*[c]: Allgaier et al., 2010. *reactor*[d]: Hollister et al., 2012. *Compost*[e]: Allgaier et al., 2010.

Discussion

Comparative analysis of taxonomic composition in the microbial consortia

The microbial consortium GF-20 was predominantly populated with Proteobacteria, followed by Bacteroidetes, Actinobacteria and Firmicutes. Proteobacteria, which is one of the largest phyla in nature, and most of the species that are related to the biomass cycle have relatively complete metabolic mechanisms and a strong resistance to stress (Guo et al.2016). Proteobacteria and Bacteroidetes can absorb compounds in the surrounding environment for growth (Siebers et al. 2018). Further analysis found that considerable differences in microbial composition of CS and FP. The microbial consortia CS primarily contained members from the genera *Pseudomonas*, *Stenotrophomonas*, *Achromobacter*, *Dysgonomonas*, *Flavobacterium* and *Sphingobacterium*, as well as *Cellvibrio*, *Azospirillum*, *Pseudomonas*, *Dysgonomonas* and *Cellulomonas* in FP.

Most microorganisms detected by metagenomic sequencing in this study were efficient biomass-degrading bacteria. However, compared with other microbial communities metagenome data, taxonomic profiling distinguished the microbial consortium GF-20 obtained from the straw returning soil ecosystem and other microbial consortia derived from soil in which sugarcane bagasse had decomposed (BGC-1) (Wongwilaiwalin et al. 2013) and a manure compost environment (EMSD5 and SAC) (Zhu et al. 2016, Allgaier et al. 2010) (Fig. 10). *Pseudomonas* was the predominant genus in CS and FP. However, *Clostridium* was extremely abundant in EMSD5 (Zhu et al. 2016) and SAC (Allgaier et al. 2010), while *Bacillus* was in SAC (Allgaier et al. 2010). This incisive taxonomic variance was more probably owing to the distribution of different physical and chemical indicators in each ecosystem. The taxonomic profiling was strongly affected by culturing conditions. Microbial consortia derived from the same inoculation source differed significantly in their composition of bacteria under different straw treatment conditions (TWS and RWS) (Jiménez et al.2015) and different carbon source conditions (SI, SI-LIG, and SI-XYL) (Carlos et al. 2018). Specifically, the microbial composition structure of GF-20 cultured under corn stover carbon sources was distinct from the filter paper source. Corn stover is composed of a matrix of heteropolymers (cellulose, hemicellulose, and lignin), and filter paper is primarily composed of cellulose. Different substrates lead to significant variation in the composition of CS and FP consortia. *Pseudomonas, Dysgonomonas, Stenotrophomonas, Sphingobacterium, Achromobacter*, and *Flavobacterium* were the predominant genera in CS. Notably, an increase in the abundance of *Pseudomonas* (rich to 63.88%) was associated with the enrichment of genes related to the degradation of aromatic compounds.

Members of the *Pseudomonas* genus are related to microorganisms that degrade aromatic compounds in various microbial growth environments (Jiménez et al.2015; Tian et al. 2016), and most of the *Pseudomonas* species can transmit signals (Marqués et al. 1993). *Dysgonomonas* species were prevalent in microbial communities from cow dung, and they played an important role in the process of degrading lignocellulosic material (Wakarchuk et al. 2016; Zhuang et al. 2015; Cherie et al. 2014). Microbes that could degrade petroleum were frequently annotated with *Stenotrophomona*, which are fibrinolytic microorganisms that can degrade alkanes (Elufisan et al. 2020). Previous studies indicated that *Sphingobacterium* and *Achromobacter* were critical microorganisms for the degradation of polycyclic aromatic hydrocarbons (PAHs), and *Sphingomonas* sp. GD542 and *A. xylosoxidans* participated in the petroleum degradation pathway and almost completely degraded C12-C23 and C27-C43 alkanes and polycyclic aromatic hydrocarbons (Carvalho et al. 2001). They also adjusted the pH value with organic acids (e.g., propionic acid, succinic acid, and L-alaninamide) during the degradation of corn stover (Duan et al. 2014). *Flavobacterium* was mostly derived from soil and can utilize the intermediate products of straw decomposition to promote its degradation (Kang et al. 2013). *Cellvibrio*, *Azospirillum*, *Cellulomonas*, and *Hydrogenophaga* were present at very low abundances in CS but were enriched in FP. The species of *Cellvibrio* listed above can secrete 3-4 types of cellulose-degrading enzymes, possess xylanase genes, and degrade carboxymethyl cellulose, cellulose, hemicellulose, and chitin (Nelson et al. 2015; Millward-Sadler et al. 1995), and *C. mixtus* is a facultative anaerobic microorganism that produces cellulase; the secreted cellulase is a typical low temperature cellulose (Wu wt al. 2015). *Azospirillum* has a synergistic effect with *Niveispirillum*, which can degrade various lignin monomers, dimers, and the natural lignin of switchgrass and alfalfa, and secrete lignin peroxidase (LiP) (Kupryashina et al. 2015). *Cellulomonas* has been recognized for its involvement in the decomposition of complex carbohydrates owing to its secretion of β-glucosidase (GH1 and GH3), endonuclease (GH6 and GH9), β-1,4-glucanase (GH48 and GH5), and cellobiose phosphorylase (GH94) that have been used in the field of biomass biodegradation and

bioremediation (Wakarchuk et al. 2016). *Hydrogenophaga* secretes oxidase and ferments oxidation products to prevent the accumulation of organic acids (Xing et al. 2018). Based on the crucial taxonomic profile of the CS and FP consortia, we hypothesize that these communities may display distinctive functional properties and may subsequently be uniquely adapted to the lignocellulosic materials in the straw returning habitat in alpine and cold regions.

Comparative analysis of lignocellulose degradation functional in the microbial consortia

The metabolic functional profile showed considerable differences of the microbial consortia CS and FP, the COG, KEGG and CAZy databases were used for annotation. The CS consortium had an increase in the acid and ester metabolism pathways such as tryptophan metabolism, glyoxylate and dicarboxylate metabolism, valine, leucine and isoleucine degradation, and butanoate metabolism, while carbohydrate metabolism was enriched in the FP consortium. In which, a two-component system present at a significantly higher abundance in CS (Fig. 5). To enhance their ability to utilize complex plant substrates, such as cellulose, lignin, and hemicellulose, and transport the resulting metabolic products, such as glucose, mannose, xylose, and ferulic acid, from various structures into the cells, the microbial consortium degrades lignocellulose by overrepresenting the genes for ABC transporters (ABCT) and two-component systems (TCS) (Zhu et al. 2015), ABCT are transmembrane proteins that utilize the energy of adenosine triphosphate hydrolysis to carry out certaun biological processes including translocation of various subtrates across membranes and non- transport-related processes such as translation of RNA and DNA repair. And TCS can coupled to environmental stimuli for an appropriate cellular response. This system senses environmental changes and regulates cellular metabolism in response to these changes thereby allowing bacteria to grow. The pH value is one of the characteristic indexes reflecting the metabolites of degradation system, the pH value of the fermentation broth of the consortia decreased first and then rose slowly under filter paper and corn stalk carbon source conditions, but the lowest pH value and the time to reach the lowest pH value were not the same (Supplementary Fig. 3). The value and time for FP treatment to reach the lowest pH value were lower and longer than those of CS treatment. It shows that the biomass of the microbial consortium increases under corn stover condition, which performs the function of straw degradation and produces organic acids, and returns to the original pH value in a short time.

Furthermore, a comparison against the CAZy database showed that the GHs and GTs genes comprise a relatively high proportion. It can be inferred that the microbial consortia CS and FP contain a rich diversity of lignocellulose-degrading bacteria and participate in the complex lignocellulose process. Among them, the proportion of microorganisms responsible for hydrolyzing glycosidic bonds between sugars or between sugars and non-sugar groups and catalyzing the transfer of glycoside bonds of various disaccharides, oligosaccharides, and polysaccharides is relatively high. Further analysis found that the microbial consortia CS and FP contain a rich diversity of lignocellulose degrading families, in which GH5, GH6, GH9, GH10, GH11, GH26, GH42, and GH43 were enriched in the FP consortium, and GH44, GH28, GH2, and GH29 increased in the CS consortium. The carbohydrate active enzyme gene of CS differs significantly from those from other lignocellulose degradative microbial communities. For example, the straw-degrading consortia RSA primarily originated from the phyla Actinobacteria, Proteobacteria, and Firmicutes (14.2%) that contain *Thermomonospora, Streptosporangium, Sphaerobacter*, and *Rhodothermus* (Wang wt al. 2016). All the GH-encoding genes in the consortium were identified as primarily being derived from *Sphingobacterium* and

Kluyvera (Jiménez et al. 2015). Notably, the functional profiles of CS and FP were similar to those of the other lignocellulose degrading microbial consortia EMSD5 and RSA (Zhu et al. 2015;Wang et al. 2016). The metagenomes of lignocellulolytic microbial consortia resembled other efficient lignocellulolytic microbial consortia, thermophilic fermenter, and rumen microbiome metagenomes (Table 3). GH5 in the FP consortia (5.56%) was slightly greater than that in the consortia RSA (2.36%) (Wang et al. 2016), BGC-1 (2.28%) (Wongwilaiwalin et al. 2013), SAC (3.2%) (Allgaier et al. 2010), reactor (3.12%) (Hollister et al. 2012) and compost (5.23%) (Allgaier et al. 2010) communities, and GH29 and GH44 in the CS consortia (2.81% and 0.88%) were higher than those of other communities. Notably, 1.47% and 2.08% of the sequences in the CS and FP microbial consortia were assigned to the GH family, respectively, and differ significantly. This indicates that the degradative mechanism of the lignocellulose mechanism of the two populations of microorganisms is similar and differed from the microbial community screened from other habitats [Allgaier et al. 2010; Wongwilaiwalin et al. 2013, Hollister et al. 2012, Hess et al. 2011].

Conclusions

This study demonstrated that the microbial consortium GF-20 cultured under conditions of complex lignocellulose (corn stover, CS) and simple cellulose (filter paper, FP) exhibits a significant degree of structural diversity. *Pseudomonas, Stenotrophomonas, Achromobacter, Dysgonomonas, Flavobacterium* and *Sphingobacterium* were extremely abundant in CS, while *Cellvibrio, Azospirillum, Pseudomonas, Acidovorax, Hydrogenophaga, Stenotrophomonas, Dysgonomonas* and *Cellulomonas* were extremely abundant in FP. Furthermore, the metabolic functions revealed considerable levels of diversity. The CS consortia had an increase in acid and ester metabolism pathways, while carbohydrate metabolism was enriched in the FP consortia. The relative abundance of genes that encode cellulases, hemicellulases, and debranching enzymes in FP was higher than those of CS, while GH3, GH2, GH29, GH78, GH28, GH38, and GH8 were significantly enriched in CS. Our findings expand the biotechnological utility of microbial populations that have been implicated in the degradation of lignocellulose. Even more importantly, they enable us to advance the understanding of the interactions between microorganisms and the synergism of enzymes that act in concert during this degradative process and guide us in the selection of different microbial consortia based on different biodegradable substrates.

Abbreviations

CS: microbial consortium GF-20 cultured on corn stover carbon media; FP: microbial consortium GF-20 cultured on filter paper carbon media; COG: Cluster of orthologous groups; KEGG: Kyoto Encyclopedia of Genes and Genomes; OTUs: Operational taxonomic units; CAZyme: Carbohydrate-active enzymes; GH: glycoside hydrolases; RA: Relative abundance. ABCT: ABC transporters; TCS: two-component systems.

Declarations

Supplementary Information

The online version contains supplementary material available at https:// .

Acknowledgements

We would like to thank the Majorbio Bio-Pharm Technology Co., Ltd. for assistance with the Illumina genome analysis.

Funding

This study was supported by the National Natural Science Foundation of China (31760353, 32060434), Inner Mongolia Natural Sciences Foundation (2020MS03086, 2018ZD02), Earmarked Fund for China Agriculture Research System (CARS-02-63), and the Crop Science Observation & Experiment Station in Loess Plateau of North China, Ministry of Agriculture, China (25204120).

Availability of data and materials

The datasets used and analyzed within this study are available from the corresponding author upon reasonable request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Authors' contributions

Qinggeer, BZ allocated funds, curated the metagenomic data, conducted data analysis, interpreted the data, drafted parts of the manuscript. XF, JL coordinated the study, collected data, interpreted data, drafted parts of the manuscript and co-wrote the manuscript. X, JW, DL,SP collected data and critically analyzed the manuscript. SC allocated funds and co-analysis the data. All authors read and approved the final manuscript

Author details

[1] Agricultural College, Inner Mongolia Agricultural University, No. 275, XinJian East Street, Hohhot 010019, China. [2] Key Laboratory of Crop Cultivation and Genetic Improvement in Inner Mongolia Autonomous Region, No. 275, XinJian East Street, Hohhot 010019, China. [3] Special Crops Institute, Inner Mongolia Academy of Agricultural & Animal Husbandry Sciences, No22, ZhaoJun Road, Hohhot, 010070, China. [4] Vocational and Technical College, Inner Mongolia Agricultural University, Baotou 014109, China. [5] Hortlculture and plant protection College, Inner Mongolia Agricultural University, No. 29, Eerduosi East Street, Hohhot 010019, China.

References

1. Allgaier M, Reddy A, Park JI, Ivanova N, Haeseleer PD', Lowry S, Sapra R, Hazen TC, Simmons BA, Vandergheynst JS (2010) Targeted discovery of glycoside hydrolases from a switchgrass-adapted compost community. Plos One, 21;5(1):e8812. https://doi.org/10.1371/journal.pone. 0008812
2. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25(17):3389–3402

3. Carlos C, Fan H, Currie CR (2018) Substrate Shift Reveals Roles for Members of Bacterial Consortia in Degradation of Plant Cell Wall Polymers. Front Microbiol 9:364
4. Carvalho MF, Vasconcelos I, Bull AT, Castro PM (2001) A GAC biofilm reactorfor the continuous degradation of 4-chlorophenol: treatment efficiency andmicrobial analysis. Appl Microbiol Biotechnol 57:419–426
5. Duan J, Liang JD, Du WJ, Wang DQ (2014) Biodegradation of Kraft Lignin by a Bacterial Strain Sphingobacterium sp. HY-H. Adv Mater Res 955–959:548–553
6. Elufisan TO, Rodríguez-Luna IC, Oyedara OO, Sánchez-Varela A, Hernández-Mendoza A, Dantán Gonzalez E, Paz-González AD, Muhammad K, Rivera G, Villalobos-Lopez MA, Guo X (2020) The Polycyclic Aromatic Hydrocarbon (PAH) degradation activities and genome analysis of a novel strain Stenotrophomonas sp. Pemsol isolated from Mexico. Peer J 8:e8102
7. Chandel AK, Singh OV (2011) Weedy lignocellulosic feedstock and microbial metabolic engineering: advancing the generation of Biofuel. Appl Microbiol Biotechnol 89(5):1289–1303
8. Chang VS, Holtzapple MT (2000) Fundamental factors affecting biomass enzymatic reactivity. Appl Biochem Biotechnol 84–86:5–37
9. Cortes-Tolalpa L, Jiménez DJ, de Lima Brossi MJ, Salles JF, van Elsas JD (2016) Different inocula produce distinctive microbial consortia with similar lignocellulose degradation capacity. Appl Microbiol Biotechnol 100(17):7713–7725
10. de Lima Brossi MJ, Jiménez DJ, Cortes-Tolalpa L, van Elsas JD (2016) Soil-Derived Microbial Consortia Enriched with Different Plant Biomass Reveal Distinct Players Acting in Lignocellulose Degradation. Microb Ecol 71(3):616–627
11. Ghose TK (1987) Measurement of cellulase activities. Pure Appl Chem 59(2):257–268
12. Guo L, Zheng S, Cao C, Li (2016) Tillage practices and straw-returning methods affect topsoil bacterial community and organic C under a rice- wheat cropping system in central China. Sci Rep 6:33155
13. Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G, Luo S, Clark DS, Chen F, Zhang T, Mackie RI, Pennacchio LA, Tringe SG, Visel A, Woyke T, Wang Z, Rubin EM (2011) Metagenomic Discovery of Biomass-Degrading Genes and Genomes from Cow Rumen. Science New Series 331:6016:463–467
14. Hollister EB, Forrest AK, Wilkinson HH, Ebbole DJ, Tringe SG, Malfatti SA, Holtzapple MT, Gentry TJ (2012) Mesophilic and thermophilic conditions select for unique but highly parallel microbial communities to perform carboxylate platform biomass conversion. PLoS One 7(6):e39689
15. Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC (2011) Integrative analysis of environmental sequences using MEGAN4. Genome Res 21(9):1552–1560
16. Jensen LJ, Julien P, Kuhn M, von Mering C, Muller J, Doerks T, Bork P (2008) eggNOG: automated construction and annotation of orthologous groups of genes. Nucleic Acids Res 36:250–254
17. Jiménez DJ, Chaves-Moreno D, van Elsas J (2015) Unveiling the metabolic potential of two soil-derived microbial consortia selected on wheat straw. Sci Rep 5:13845
18. Jiménez DJ, Maruthamuthu M, van Elsas JD (2015) Metasecretome analysis of a lignocellulolytic microbial consortium grown on wheat straw, xylan and xylose. Biotechnol Biofuels 8:199

19. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res, 1;28(1):27-30. https:// doi.org/10.1093/nar/28.1.27
20. Kang JY, Chun J, Jahng KY (2013) Flavobacterium aciduliphilum sp. nov. isolated from freshwater, and emended description of the genus Flavobacterium. Int J Syst Evol Microbiol 63(5):1633–1638
21. Kupryashina MA, Petrov SV, Ponomareva EG, Nikitina VE (2015) Ligninolytic activity of bacteria of the genera *Azospirillum* and *Niveispirillum. Mikrobiologiia*, 84(6):691-696. https:// doi.org/10.1134/S0026261715060041
22. Lemos LN, Pereira RV, Quaggio RB, Martins LF, Moura LMS, da Silva AR (2017) Genome-centric analysis of a thermophilic andcellulolytic bacterial consortium derived from composting. Front Microbiol 8:644
23. Lewin GR, Johnson AL, Soto RD, Perry K, Book AJ, Horn HA, Pinto-Tomás AA, Currie CR (2016) Cellulose-Enriched Microbial Communities from Leaf-Cutter Ant (Atta colombica) Refuse Dumps Vary in Taxonomic Composition and Degradation Ability. PLoS One, 21;11(3):e0151840. https://doi.org/10.1371/journal.pone.0151840
24. Limayem A, Ricke SC (2012) Lignocellulosic biomass for bioethanol production: Current perspectives, potential issues and future prospects. Progress in Energy & Combustion Science 38(4):449–467
25. Noguchi H, Park J, Takagi T (2006) MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. Nucleic Acids Res 34(19):5623–5630
26. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:490–5
27. López-Mondéjar R, Zühlke D, Becher D, Riedel K, Baldrian P (2016) Cellulose and hemicellulose decomposition by forest soil bacteria proceeds by the action of structurally variable enzymatic systems. Sci Rep 6:25279
28. Maruthamuthu M, Jiménez DJ, Stevens P, van Elsas JD (2016) A multi-substrate approach for functional metagenomics-based screening for (hemi)cellulases in two wheat straw-degrading microbial consortia unveils novel thermoalkaliphilic enzymes. BMC Genomics 17:86
29. Marqués S, Ramos JL (1993) Transcriptional control of the Pseudomonas putida TOL plasmid catabolic pathways. Mol Microbiol 9(5):923–935
30. Minty JJ, Singer ME, Scholz SA, Bae CH, Ahn JH, Foster CE, Liao JC, Lin XN (2013) Design and characterization of synthetic fungal-bacterial consortia for direct production of isobutanol from cellulosic biomass. Proc Natl Acad Sci U S A 110(36):14592–14597
31. Millward-Sadler SJ, Davidson K, Hazlewood GP, Black GW, Gilbert HJ, Clarke JH (1995) Novel cellulose-binding domains, NodB homologues and conserved modular architecture in xylanases from the aerobic soil bacteria Pseudomonas fluorescens subsp. cellulosa and Cellvibrio mixtus. Biochem J 1995 312:39–48
32. Nelson CE, Gardner JG, Gardner (2015) In-Frame Deletions Allow Functional Characterization of Complex Cellulose Degradation Phenotypes in Cellvibrio japonicas. Appl Environ Microbiol 81(17):5968–5975
33. Parks DH, Beiko RG (2010) Identifying biologically relevant differences between metagenomic communities. Bioinformatics 26(6):715–721
34. Qinggeer, Gao JL, Yu XF, Zhang BL, Wang ZG, Borjigin N, Hu SP, Sun JY, Xie M, Wang Z (2016) Screening of a microbial consortium with efficient corn stover degradation ability at low temperature. J.Integr.Agr. 2016,

15:2369-2379. https://doi org/10.1016/S2095-3119(15)61272-2

35. Sheng P, Huang J, Zhang Z, Wang D, Tian X, Ding J (2016) Construction and characterization of a cellulolytic consortium enriched from the hindgut of holotrichia parallela Larvae. Int J Mol Sci 17(10):1646
36. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA (2003) The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4:41
37. Siebers M, Rohr T, Ventura M, Schütz V, Thies S, Kovacic F, Jaeger KE, Berg M, Dörmann P, Schulz M (2018) Disruption of microbial community composition and identification of plant growth promoting microorganisms after exposure of soil to rapeseed-derived glucosinolates. PLoS One 13(7):e0200160
38. Tian JH, Pourcher AM, Peu P (2016) Isolation of bacterial strains able to metabolize lignin and lignin-related compounds. Lett Appl Microbiol 63:30–37
39. Wakarchuk WW, Brochu D, Foote S, Robotham A, Saxena H, Erak T, Kelly J (2016) Proteomic analysis of the secretome of *Cellulomonas fimi* ATCC 484 and *Cellulomonas flavigena* ATCC 482. PLoS ONE 11(3):e0151186
40. Wang C, Dong D, Wang H, Müller K, Qin Y, Wang H, Wu W (2016) Metagenomic analysis of microbial consortia enriched from compost: new insights into the role of Actinobacteria in lignocellulose decomposition. Biotechnol Biofuels 9:22
41. Wongwilaiwalin S, Laothanachareon T, Mhuantong W, Tangphatsornruang S, Eurwilaichitr L, Igarashi Y, Champreda V (2013) Comparative metagenomic analysis of microcosm structures and lignocellulolytic enzyme systems of symbiotic biomass-degrading consortia. Appl Microbiol Biotechnol 97(20):8941–8954
42. Wu YR, He J (2015) Characterization of a xylanase-producing Cellvibrio mixtus strain J3-8 and its genome analysis. Sci Rep 5:10521
43. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li CY, Wei L (2011) KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases.Nucleic Acids Res,W316-W322. https:// doi.org/10.1093/nar/gkr483
44. Xing W, Li J, Li P, Wang C, Cao Y, Li D, Yang Y, Zhou J, Zuo J (2018) Effects of residual organics in municipal wastewater on hydrogenotrophic denitrifying microbial communities. J Environ Sci 65(3):262–270
45. Yu XF, Qinggeer, Gao JL, Wang ZG, Hu SP, Borjigin N, Wang Z, Sun JY, Han SC (2019) Exploration of the key microbes and composition stability of microbial consortium GF-20 with efficiently decomposes corn stover at low temperatures. J.Integr.Agr, 2019,8:1893-1904. https://doi.org/10.1016/S2095-3119(19)62609-2
46. Zhuang W, Zhang S, Xia X, Wang G (2015) Draft genome sequence of Cellulomonas carbonis T26 T and comparative analysis of six Cellulomonas genomes. Standards in Genomic Sciences 10(1):104
47. Zhu N, Yang J, Ji L, Liu J, Yang Y, Yuan H (2016) Metagenomic and metaproteomic analyses of a corn stover-adapted microbial consortium EMSD5 reveal its taxonomic and enzymatic basis for degrading lignocellulose. Biotechnol Biofuels 9(1):243
48. Ziemer CJ (2014) Newly cultured bacteria with broad diversity isolated from eight-week continuous culture enrichments of cow feces on complex polysaccharides. Appl Environ Microbiol 80(2):574–585
49. Table **1** Summary of Illumina GA reads and ORF prediction

Figures

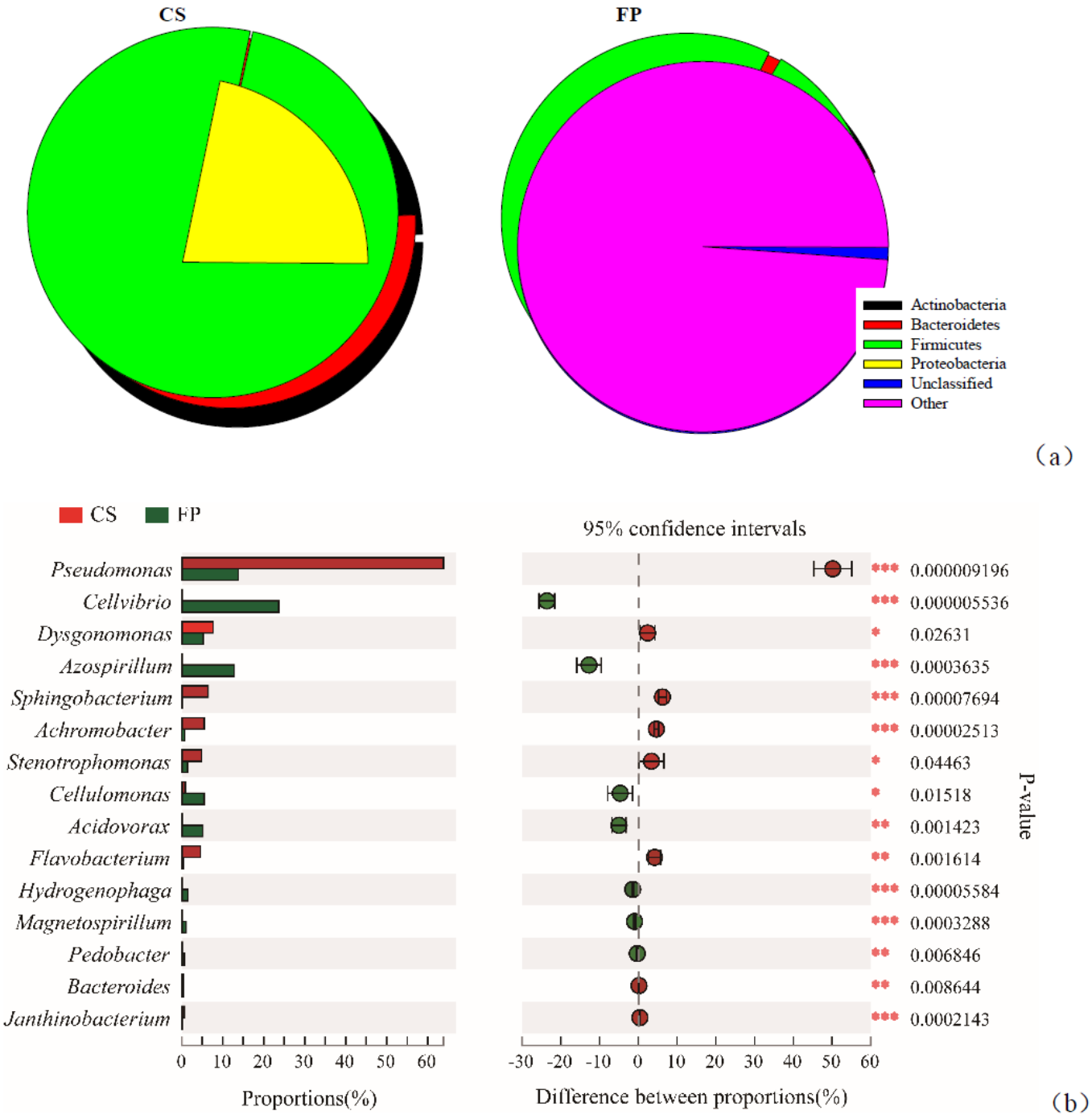
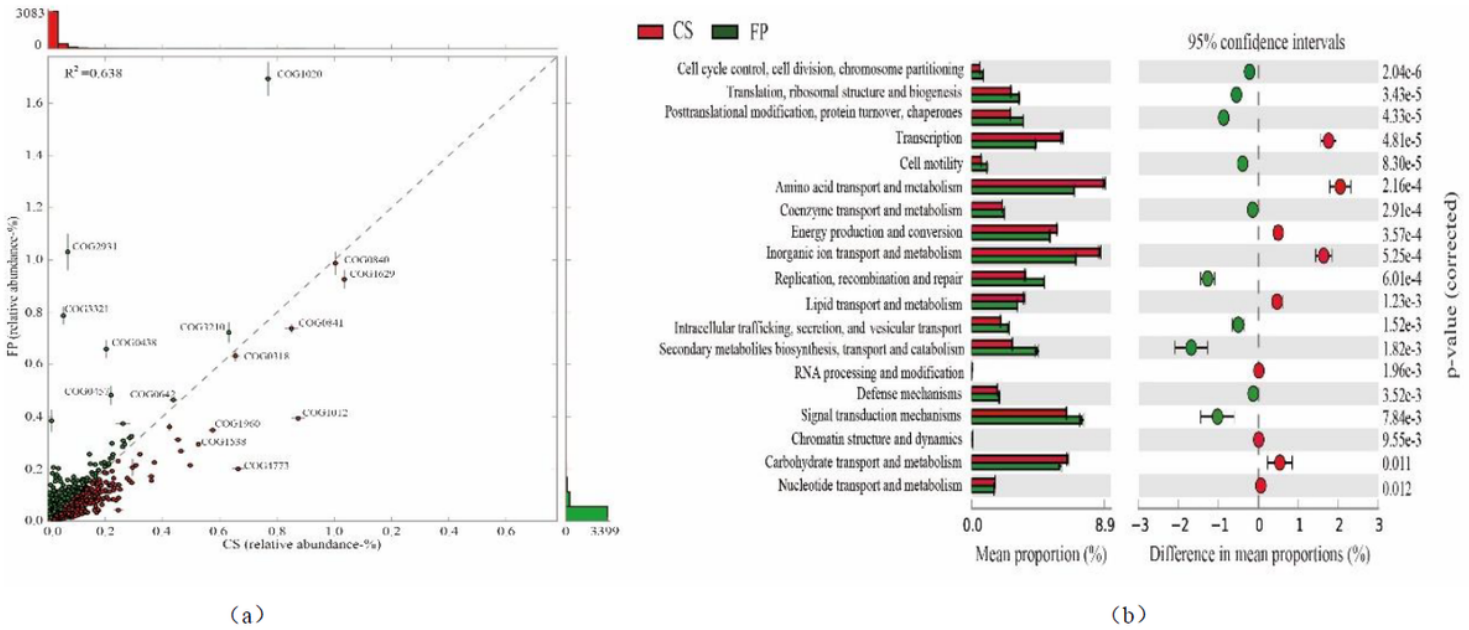


Figure 1

Relative abundances of phyla (a) and differentially dominant genera (p<0.05,95% CIs) (b) in each microbial consortium Note: The phyla with a relative abundance lower than 0.1% were assigned as other, and the sequences with a classification confidence threshold in a lineage without information in the NR databases were listed as unclassified



(a) (b)

Figure 2

Functional profile of the metagenomes based on COG identifiers (a) Comparison of functional profiles at the COG function between the CS and FP consortia. (b) Differentially enriched functions ($p<0.05$, 95% CIs) at the COG function between the CS and FP consortia.

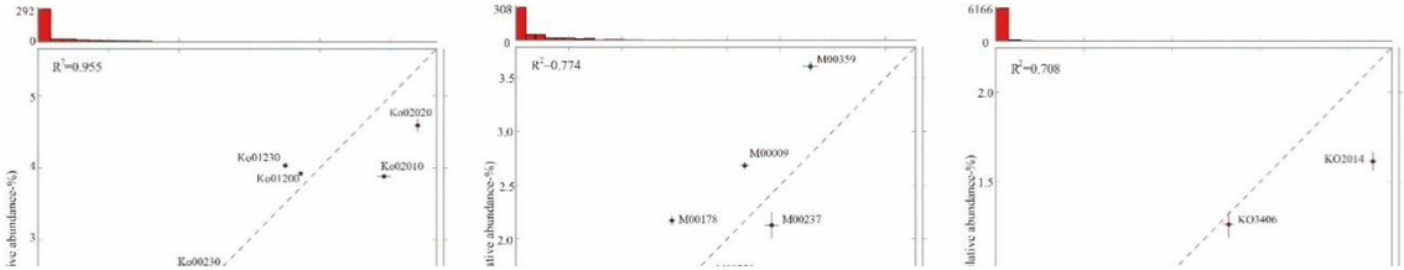


Figure 3

Comparison of the functional profiles at pathways(a), modules(b) and KEGG Orthologys(c) between the microbial consortia CS and FP.

Figure 4

Differentially enriched functions (p<0.05, 95% CIs) in KEGG at level two between the microbial consortia CS and FP.

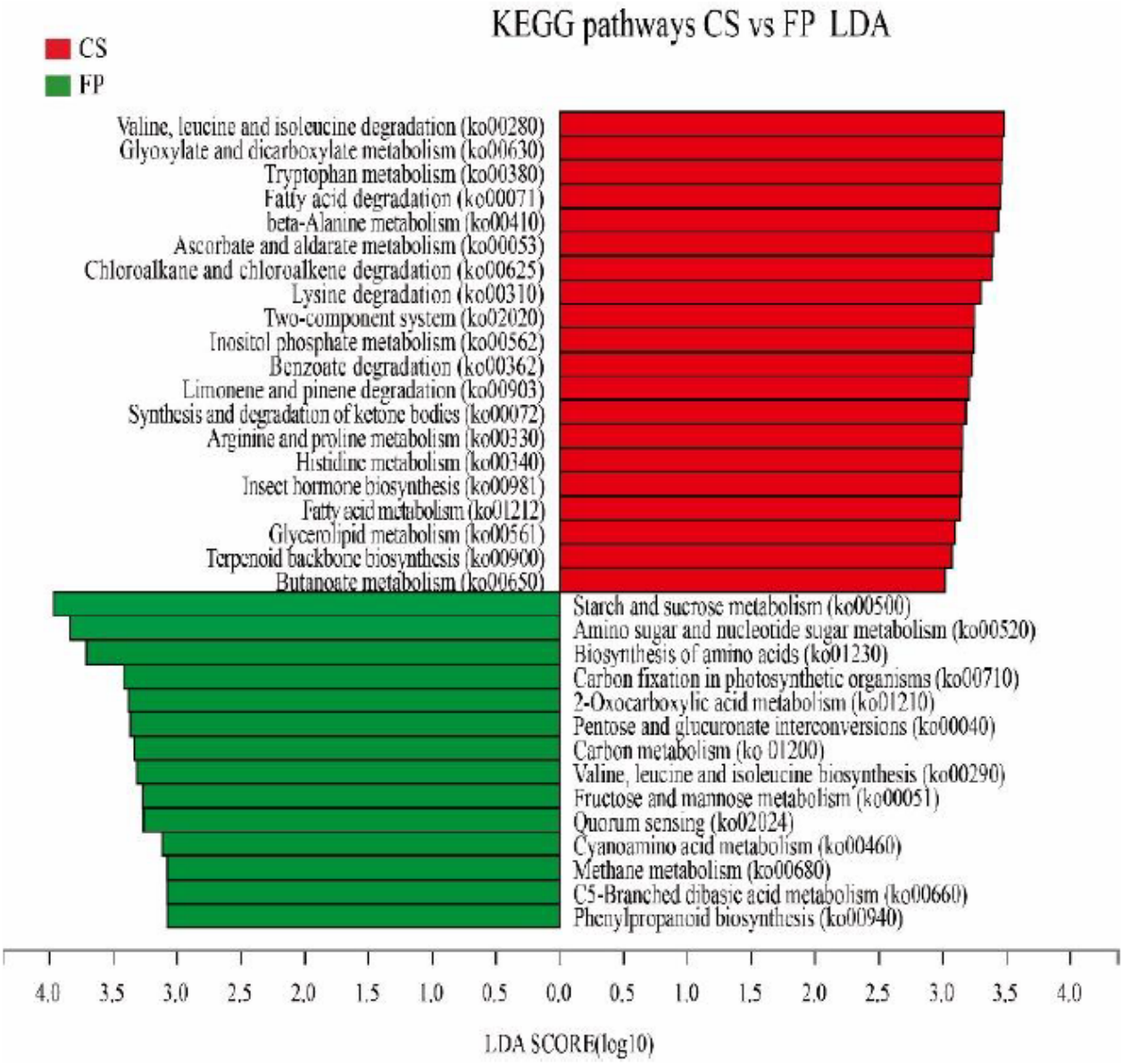


Figure 5

Linear discriminative analysis (LDA) effect size Lefse analysis of KEGG pathways at carbohydrate metabolism level

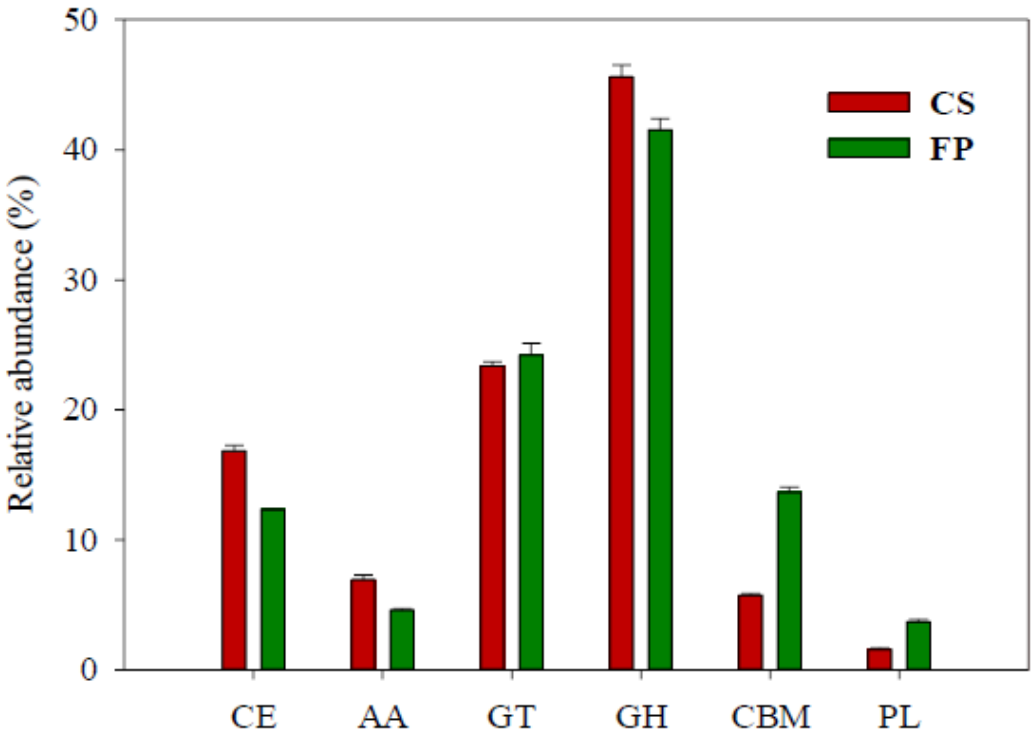


Figure 6

Distribution of predicted reads in the CAZymes at the class level between the microbial consortia CS and FP. CBM: carbohydrate binding modules, CE: carbohydrate esterases, GH: glycosyl hydrolases, GT: glycosyl transferases, PL: polysaccharide lyases.

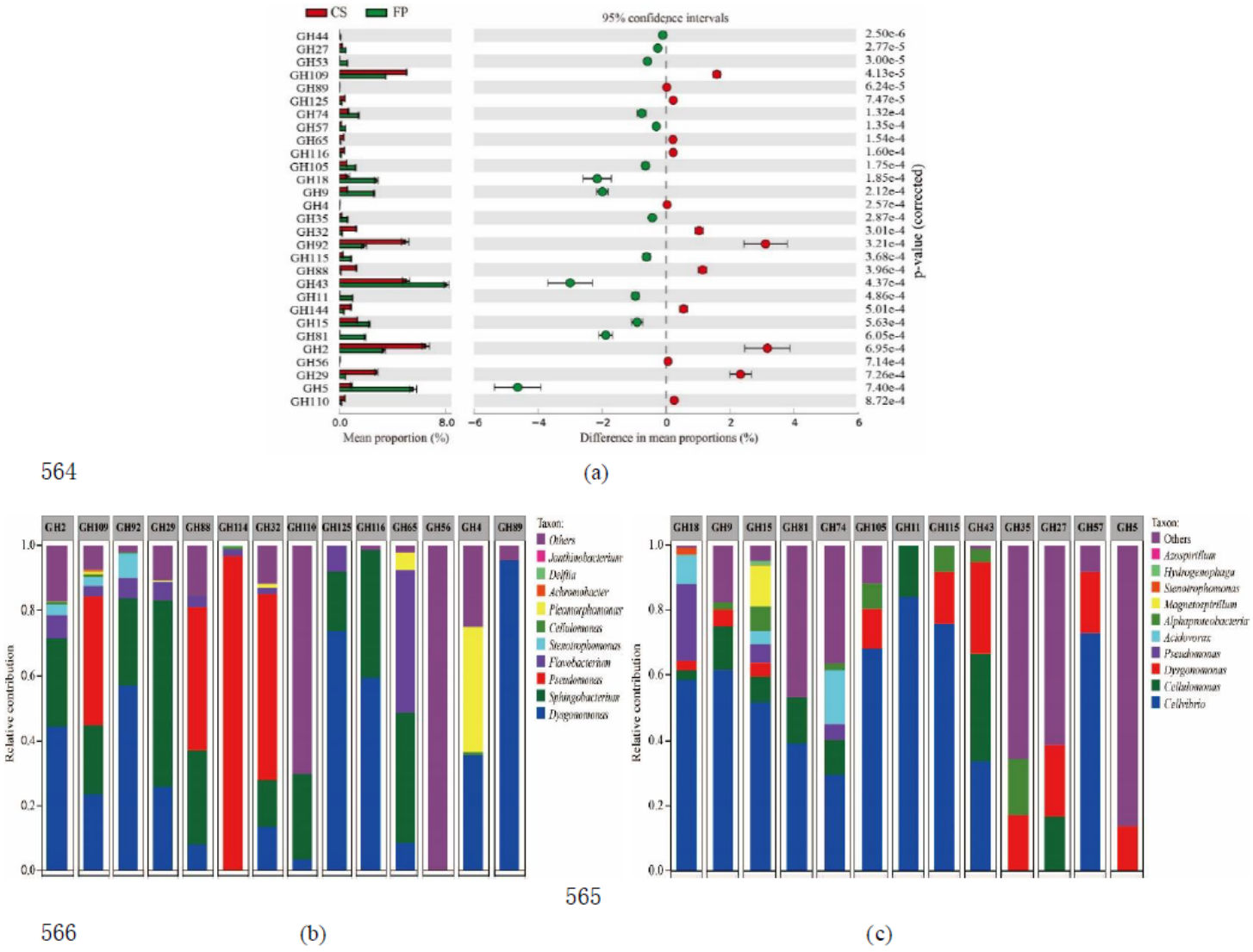


Figure 7

Differentially enriched glycoside hydrolase families ($p < 0.001$, 95% confidence intervals) between CS and FP (a) metagenomes; (b) Taxonomic affiliation of reads that belong to the families GH2, GH109, GH92, GH29, GH88, GH114, GH110, GH32, GH125, GH116, GH65, GH56, GH4, and GH89 in CS; (c) Taxonomic affiliation of reads that belong to the families GH18, GH9, GH15, GH81, GH74, GH105, GH11, GH115, GH43, GH35, GH27, GH57, and GH5 in FP using the Lowest Common Ancestor (LCA) algorithm.

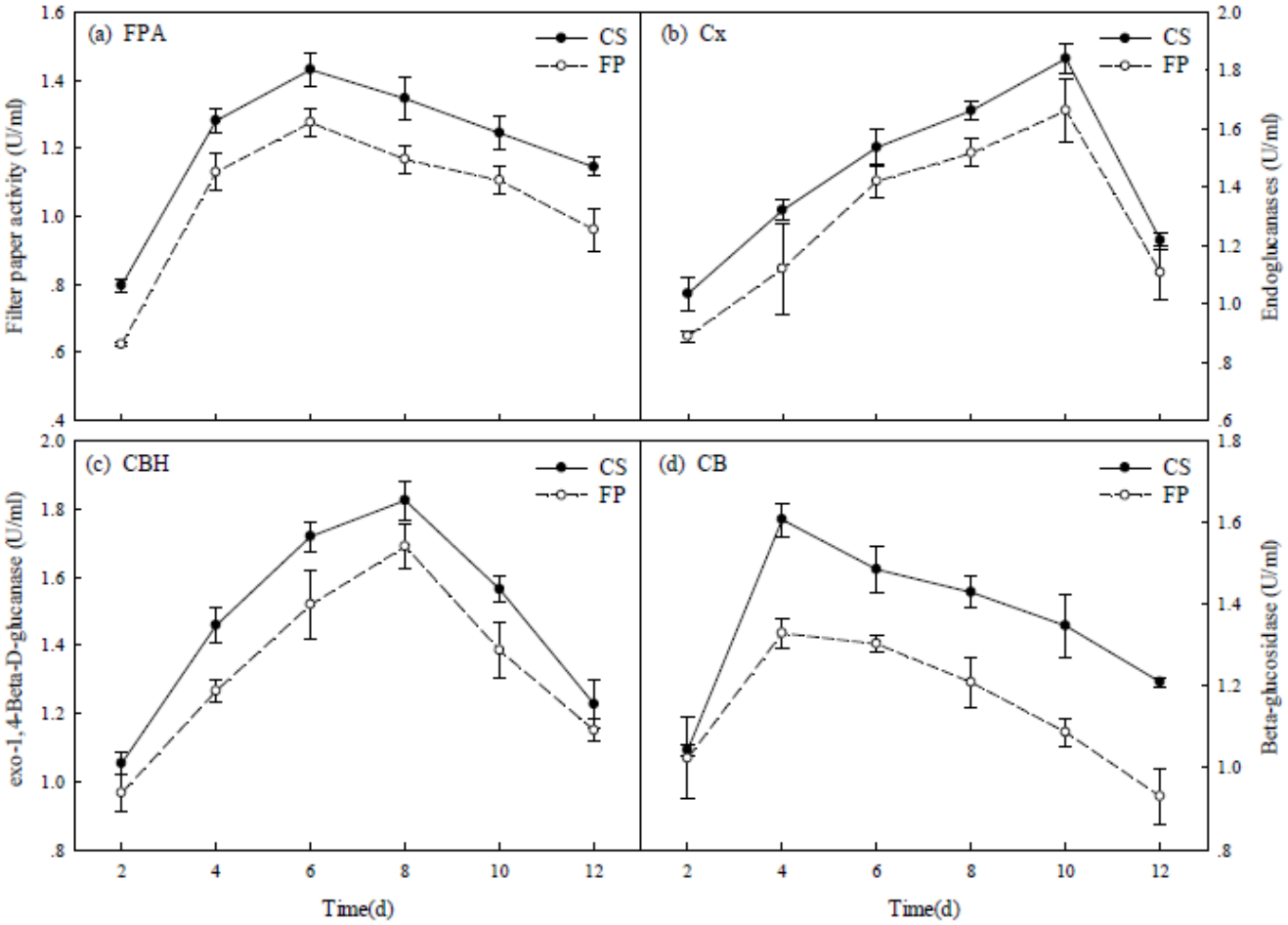


Figure 8

Dynamics of cellulase enzyme production by microbial consortium GF-20 under corn stover(CS) and filter paper (FP) carbon conditions.

Figure 9

Taxonomic affiliation of the reads that belong to the families GH3, GH2, GH8, GH28, GH38, GH78, and GH29 in CS using the Lowest Common Ancestor (LCA) algorithm.

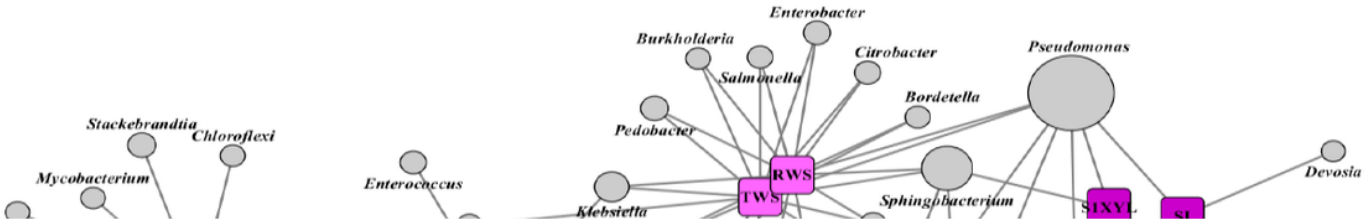


Figure 10

Collinear network diagram for metagenome taxonomic profiling in the genus level from other consortia samples, as well as in GF-20 (CS and FP) Note：TWS,RWS：Jiménez et al., 2015；SI：SIXYL,SILIG: Carlos et al., 2018, SAC: Allgaier et al., 2010, EMSD5:Zhu et al., 2016, BGC-1:Wongwilaiwalin et al.,2013.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplemenrtyFigure1.png
- SupplemenrtyFigure2.png
- SupplemenrtyFigure3.png
- qinggeerAuthorchecklist.docx