Metagenomic Insights Into a Cellulose-Rich Niche Reveal Microbial Cooperation in Cellulose Degradation

Jinming Cui[1,2,3†], Guoqin Mai[1,3†], Zuowei Wang[2,3], Quan Liu[1,3], Yan Zhou[1,3], Yingfei Ma[1,3*] and Chenli Liu[1,2,3*]

[1] Institute of Synthetic Biology – Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, [2] Guangzhou Institute of Advanced Technology, Chinese Academy of Sciences, Guangzhou, China, [3] Shenzhen Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

Background: Cellulose is the most abundant organic polymer mainly produced by plants in nature. It is insoluble and highly resistant to enzymatic hydrolysis. Cellulolytic microorganisms that are capable of producing a battery of related enzymes play an important role in recycling cellulose-rich plant biomass. Effective cellulose degradation by multiple synergic microorganisms has been observed within a defined microbial consortium in the lab culture. Metagenomic analysis may enable us to understand how microbes cooperate in cellulose degradation in a more complex microbial free-living ecosystem in nature.

Results: Here we investigated a typical cellulose-rich and alkaline niche where constituent microbes survive through inter-genera cooperation in cellulose utilization. The niche has been generated in an ancient paper-making plant, which has served as an isolated habitat for over 7 centuries. Combined amplicon-based sequencing of 16S rRNA genes and metagenomic sequencing, our analyses showed a microbial composition with 6 dominant genera including *Cloacibacterium*, *Paludibacter*, *Exiguobacterium*, *Acetivibrio*, *Tolumonas*, and *Clostridium* in this cellulose-rich niche; the composition is distinct from other cellulose-rich niches including a modern paper mill, bamboo soil, wild giant panda guts, and termite hindguts. In total, 11,676 genes of 96 glucoside hydrolase (GH) families, as well as 1,744 genes of carbohydrate transporters were identified, and modeling analysis of two representative genes suggested that these glucoside hydrolases likely evolved to adapt to alkaline environments. Further reconstruction of the microbial draft genomes by binning the assembled contigs predicted a mutualistic interaction between the dominant microbes regarding the cellulolytic process in the niche, with *Paludibacter* and *Clostridium* acting as helpers that produce endoglucanases, and *Cloacibacterium, Exiguobacterium, Acetivibrio,* and *Tolumonas* being beneficiaries that cross-feed on the cellodextrins by oligosaccharide uptake.

Conclusion: The analysis of the key genes involved in cellulose degradation and reconstruction of the microbial draft genomes by binning the assembled contigs

predicted a mutualistic interaction based on public goods regarding the cellulolytic process in the niche, suggesting that in the studied microbial consortium, free-living bacteria likely survive on each other by acquisition and exchange of metabolites. Knowledge gained from this study will facilitate the design of complex microbial communities with a better performance in industrial bioprocesses.

Keywords: metagenomics, mutualistic interaction, cellulolytic process, cellulose-degrading microbial community, genomic sequencing

INTRODUCTION

Cellulose, the main component of the cell wall of green plants, algae, and oomycetes, is the most abundant organic polymer on Earth, and has an undeniably central role in the global carbon cycle (Klemm et al., 2005). Consequently, cellulolytic processes occur widely in the Earth's biosphere. In these processes, microbes employ cellulases to break down cellulose into oligosaccharides called cellodextrins or completely into glucose units, followed by sugar uptake and assimilation.

With the cellulolytic process providing a carbon and energy source, microbes proliferate and many microbial ecosystems evolve. For example, symbiotic microorganisms in the guts help some animals, particularly ruminants, giant pandas, and termites, digest cellulose; also, microbial consortia constantly recycle the cellulose-rich plant biomass in soil environments. A few examples of these ecosystems have been investigated to understand their function and dynamics (Warnecke et al., 2007; Zhu et al., 2011; He et al., 2013; Zhou et al., 2014; Lin et al., 2015).

In addition to natural recycling, cellulose has historically been used to produce paper. A fascinating example lies in an ancient village named Dengcun in Guangdong Province, China, where people have been using local bamboo and water resources to make paper continuously for 7 centuries. The centuries-long papermaking process has generated a special cellulose-rich and alkaline niche, and presumably a particular microbial ecosystem has been evolving in situ. Two intriguing questions can be raised: (i) what is the composition of the microbial ecosystem; (ii) what roles are these microbes playing in the cellulolytic process?

Cellulose is a sustainable carbon and energy source, and thus cellulolytic microbes and their enzymes can play important roles in industrial applications, such as biofuel production (Carroll and Somerville, 2009). Many studies have focused on single cellulolytic species based on a culture-dependent method, such as Clostridium thermocellum, C. cellulolyticum, and Caldicellulosiruptor bescii, most of which belong to the phylum Firmicutes (Demain et al., 2005; Dash et al., 2016). Nevertheless, the natural cellulolytic process is usually interplay within a microbial consortium. According to the BQH, in a given ecosystem, the evolution of the microbial

community tends to make the constituent microbes rely on each other (Morris et al., 2012). Microbial cells (or species) can be considered as players competing and cooperating in situ, and they likely would have employed an ESS and become highly efficient in, for example, cellulose degradation (Hummert et al., 2014). In this study, we aimed to reveal the microbial composition and adaption mechanisms of the special cellulolytic ecosystem in Dengcun. Metagenomic data were analyzed and compared with the results from other samples rich in bamboo, i.e., bamboo soil (Lin et al., 2015) and wild giant panda guts (Zhu et al., 2011). In addition, modern paper mill (Lu and Lu, 2014) and termite hindguts (He et al., 2013) samples were also included for comparison. New cellulolytic species, as well as the ESS revealed by the culture-independent metagenomic analysis method in this study, can improve the future design of synthetic microbial consortia for cellulose conversion (Kudo et al., 1987; Kato et al., 2004, 2005; Salimi et al., 2010).

MATERIALS AND METHODS

Sample Collection

Dengcun is located at longitude 112.588° E, latitude 23.3268° N, and at an altitude of 67.5 m. APMP samples 1 and 2 were collected from a bamboo fiber soaking pit (water depth 30 cm), and a bamboo pulp pit (water depth 120 cm), respectively, of an ancient paper-making plant in the biggest local workshop. After stirring, 1.5 L of liquid/pulp mixtures of each sample were collected in centrifuge bottles and shipped at 4°C, then frozen and stored at −80°C. Water quality and parameters are listed in **Supplementary Table S1**. The concentrations of total phosphorus (TP), total nitrogen (TN), dissolved organic carbon (DOC), ammonium (NH4-N), nitrate (NO3-N), nitrite (NO2-N), and soluble phosphorus (PO4-P) were measured using standard methods (Gilcreas, 1967).

To extract bacterial DNA, samples were mixed using a vortex for 5 min, followed by centrifugation at 100 g for 10 min to separate the bamboo fibers. Fifty mL of the supernatants were spun down at 4,600 g for 30 min, and the precipitates were washed twice with 10 mM pH 7.4 potassium phosphate buffer. In both samples, the pellets appeared blackish, and pea-sized samples were taken and the bacterial genomic DNA was extracted using the Rapid Soil DNA Isolation Kit by Sangon Biotech Co., Ltd., (Shanghai). Clear bands of genomic DNA were observed using polyacrylamide gel electrophoresis.

Abbreviations: APMP, ancient paper-making pulp; BQH, black queen hypothesis; ESS, evolutionary stable strategy; GH, glycoside hydrolase; LBPSDB, latest bacterial protein sequence database; OTU, operational taxonomic unit; PCA, principal component analysis; TCDB, transporter classification database.

16S rRNA Gene Amplicon-Based Sequencing

Briefly, primer pairs targeting the 16S rRNA gene v3-v4 region (515–806) were employed for amplicon-based sequencing on total DNA of samples 1 and 2, respectively. The F515 primer (5′TATGGTAATTGTGTGCCAGCMGCCGCGGTA A3′) was used across all samples. We added a linker and two barcode sequences at the 5′ end of R803 primer (5′AGTCAGTCAGCCGGACTACHVGGGTWTCTAAT3′), for sequencing of pooled samples (**Supplementary Table S2**). PCRs were set up containing 2 μl of APMP genomic DNA as template, 2 μl of F515 primer, 2 μl of R803 primer, 4 μl of dNTP, 4 μl of 25 mmol l^{-1} MgCl2, 5 μl of 10 × Ex Taq buffer, 0.25 μl Taq polymerase (5 U μl^{-1}), and 30.75 μl of distilled deionized water. PCR began with 98°C 1 min, followed by 30 cycles of 98°C 10 s, 58°C 30 s, 72°C 2 min, and finally 72°C 10 min. The PCR products were purified and preserved in 25 μl sterile water using the San-Prep column DNA gel extraction kit (Sangon Biotech, Shanhai, China). Concentrations of the purified PCR products were determined using NanoDrop (Thermo fisher Scientific Inc. Wltham, MA, United States); 200 ng of the purified PCR products of each sample were pooled together equally and then were sent to Novogene Genomics Co., Ltd., (Beijing) for Illumina Mi-Seq PE250 platform. The generated reads were firstly filtered by quality (Q > 30), and the high quality paired-end reads were assembled into a complete sequence with a minimum overlap of 50 bp using the software Qiime (Caporaso et al., 2010). Sequences shorter than 300 bp with an expected error of more than 0.5 were discarded. Chimera checking of the sequences was performed using the software Qiime. Sequences were assigned to samples according to the barcodes. Barcodes and primers were removed from the sequences for further analysis. The sequencing data have been deposited in NCBI GenBank with the accession numbers SAMN08818942.

Taxonomic Classification of 16S rRNA Gene Sequences

16S rRNA gene amplicon-based sequencing reads were assigned to OTUs at genus and phylum levels according to the method described by Ravel et al. (2011). Briefly, the RDPII classifier was employed (Wang et al., 2007): if the median RDP confidence value of a 16S rRNA gene sequence assigned to a genus was less than the 0.5 thresholds, the median RDP score for the next-higher taxonomic level (family) was considered. If it was above the RDP confidence value threshold (0.5) then the OTU was named as "FamilyName_genus-name." For each taxonomic level, the algorithm was applied in an iterative manner until the RDP confidence value was above 0.5. A shell script was employed to count the number of sequences for each genus in each sample. Genera that only appeared in one sample were discarded. PCA was performed by the software of PAST (version 3.20) based on the relative abundance (Hammer, 2001).

Published sequencing data of the samples from a modern paper mill (Lu and Lu, 2014), bamboo soil (Lin et al., 2015), wild giant panda guts (Zhu et al., 2011), and termite hindguts (He et al., 2013) are listed in **Supplementary Table S3** and were also analyzed in this study using the same bioinformatic pipelines.

Metagenomic Sequencing, *de novo* Assembly, Gene Prediction, and Annotation

Sixty μL of 300–500 ng/μL DNA of samples 1 and 2 were submitted for metagenomic sequencing to Berry Genomics Co., Ltd., (Beijing). Whole metagenomic shotgun sequencing was performed using the Illumina Hi-Seq2500 V4 PE125 (Paired-end 125 bp) platform. Briefly, whole bacterial genomic DNA was treated using a next-generation sequencing Fast DNA Library Prep Set for Illumina Hi-Seq 2500 V4 PE125 (Paired-end 125 bp) to prepare the metagenomic library following a standard Illumina protocol. The retrieved DNA library was evaluated by the StepOnePlus Real-Time PCR system (Applied Biosystems), and considered to be qualified showing a single peak with a concentration above 3 nM and a volume above 25 μL. The generated sequencing reads were firstly filtered by quality. Low-quality reads (reads containing adaptor sequence, reads with $N > 3\%$, reads where $\geq 50\%$ of the bases had a Q score ≤ 30) were removed. Clean reads were assembled by MetaVelvet (Namiki et al., 2012) with the default parameter of k-mer size of 39. Contigs larger than 300 bp were kept for further analysis. Statistics of the sequencing and assembly data were listed in **Supplementary Table S4**, and the data have been deposited in NCBI GenBank with the accession numbers SAMN07807411 for sample 1 and SAMN07807412 for sample 2.

Genes encoded by the contigs (≥ 300 bp) were predicted by MetaGeneMark (Tang et al., 2013). The clusters of orthologous groups (COG) database was downloaded from NCBI and was formatted as local COG database[1] using the program Makeblastdb (Galperin et al., 2015). All protein sequences derived from the NCBI bacterial genomes were formatted as a local database (Latest Bacterial protein sequence database, LBPSDB) using the program Makeblastdb[2]. The predicted genes were further annotated functionally by the program Blastp against the COG database and the LBPSDB (Galperin et al., 2015) with cutoff e-value 1e-10. The taxonomic affiliation of each gene was assigned at the genus level according to the best-hit counterpart in the LBPSDB. The number of the APMP predicted proteins of each genus was counted, and the mean of identities and the standard deviation (SD) between the APMP protein sequences and their counterparts in the LBPSDB database were calculated, respectively (**Supplementary Table S5**). In addition, the predicted protein sequences corresponding to the top abundant genera in APMP samples were extracted, respectively and then were uploaded to KEGG mapper[3] (Kanehisa et al., 2016), to identify cellulose-degrading enzymes and cellodextrin transporters in the reconstructed KEGG pathways (**Supplementary Table S5**).

---

[1]https://www.ncbi.nlm.nih.gov/COG/(September 22, 2016)

[2]ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/(January 09, 2016)

[3]http://www.kegg.jp/kegg/tool/annotate_sequence.html

Cellulase Analysis

With reference to the method of this article (Warnecke et al., 2007), the predicted protein sequences encoded by the contigs (≥300 bp) were compared against the dbCAN database[4] by HMMER 3.0 hmmscan with cutoff evalue 1e-10 (Warnecke et al., 2007). The dbCAN database is a knowledge-based resource specialized in the enzymes that bind, synthesize, or break down complex carbohydrates (Lombard et al., 2014a). The predicted proteins with functional domains of GH and carbohydrate binding modules (CBM) were assigned as putative cellulase for further analysis. GHs and CBMs were named according to the carbohydrate-active enzyme (CAZy) nomenclature scheme (Cantarel et al., 2009; Lombard et al., 2014b). The genes of 13 GH families that encode cellulases, namely GH5-9, GH12, GH44, GH45, GH48, GH51, GH61, GH74, and GH124 (Sukharnikov et al., 2011), were further verified by being compared with homologs in the non-redundant protein database of Genbank by BLASTP. The number of the APMP genes in each GH family was given in **Supplementary Table S6**.

All full-length genes of the 13 GH families were listed in **Supplementary Table S7**, and functional domains, including GH and CBM, were also recognized with the software HMMER 3.0 hmmscan. Gene-40972 of the GH5 family and gene-6561 of the GH9 family were taken as examples, due to their high homologies with their counterparts in the RCSB PDB database[5]. The amino acid compositions, charges, and isoelectric points of gene-40972 and gene-6561, together with their homologs, were analyzed by EMBOSS Pepstats (Li et al., 2015). The protein structures of gene-40972 and gene-6561 were predicted using the Galaxy template-based modeling program (Shin et al., 2014). *C. thermocellum* CtCelC (pdb ID: 1ceo) was used as a template for gene-40972 (Dominguez et al., 1996); *Alicyclobacillus acidocaldarius* AaCel9A (pdb ID: 3gzk) and leaf-branch compost LC-CelG (pdb ID: 3x17) were used as templates for gene-6561 (Eckert et al., 2009; Okano et al., 2015). The ligand cellotriose was predicted and docked into the models by GalaxySite (Shin et al., 2014). All structure figures were prepared using the PyMOL molecular graphics system.

Carbohydrate Transporters Analysis

All genes predicted from the APMP metagenomic contigs (≥300 bp) were compared against the TCDB[6] by the program HMMER 3.0 hmmscan with an E-value 1e-10 to identify the genes encoding carbohydrate transporters. The genes with hits were counted and the corresponding protein sequences were selected as carbohydrate transporters for further analyses. The figures were drawn by IBS (Liu et al., 2015).

Binning Metagenomic Contigs by Vizbin

Reference-independent approaches were sought to cluster the metagenomic contigs (binning) by exploiting the inherent genomic signatures. In order to get high coverage of the bacterial genomes, the contigs of two APMP samples were combined for binning analysis and the larger contigs (≥1000 bp)

were clustered by Vizbin (Laczny et al., 2015)with the default parameters: Minimal contig length 1000, Number of threads 1, Kmer length 5, Intermediate dimensions 50, Theta 0.5, Perplexity 30.0, seed 0. The coding genes of these contigs were analyzed by AmphoraNet for searching phylogenetic marker genes to estimate the taxonomic composition of the bacterial community of the metagenomes[7] (**Supplementary Table S8**; Kerepesi et al., 2014). The contigs carrying the marker genes of the top abundant genera were marked using different colored stars to show the taxonomy of each cluster where the corresponding marker genes exist. Meanwhile, the contigs carrying the genes encoding endoglucanase, beta glucosidase, 6-phospho-beta-glucosidase, and cellobiose phosphorylase, were marked using different patterned stars to show the function of the phylotype represented by each cluster.

RESULTS

Profiling of the Microbial Consortium in the Ancient Paper-Making Pulp (APMP) Samples

We profiled the microbial consortium inhabiting the APMP ecosystem by analysis of the 16S rRNA genes. In total, 15470 raw reads were obtained. After quality filter and assembly of the paired-end reads, 2581 and 3486 sequences were generated in APMP samples 1 and 2, respectively (**Supplementary Figure S1**). In total, 8 and 17 phyla, and 207 and 331 genera were identified in APMP samples 1 and 2, respectively. On average, 98.21% of the total reads were assigned to 4 phyla, namely Bacteroidetes (42.44%), Firmicutes (30.79%), Proteobacteria (23.58%), and Actinobacteria (1.40%) in both APMP samples (**Figure 1A**). The most abundant genera (in samples 1 and 2, respectively) were *Cloacibacterium* (27.82/11.91%), *Paludibacter* (8.29/34.20%), *Exiguobacterium* (31.08/11.01%), *Tolumonas* (4.18/2.79 %), and *Aeromonas* (7.81/27.02%) (**Figure 1B**). Members of the five genera are likely obligate or facultative anaerobes.

The microbial compositions were distinct among these samples, with the most abundant phylum being Bacteroidetes (33.43%) in the modern paper mill sample (Lu and Lu, 2014); Acidobacteria (37.76%) in the bamboo-soil sample (Lin et al., 2015); Firmicutes (56.97%) in the bamboo-feeding wild giant panda gut sample (Zhu et al., 2011); and Spirochaetes (56.61%) in the wood-feeding termite hindgut sample (He et al., 2013). PCA analysis of the genus distribution data indicated that the microbial communities of APMP samples 1 and 2 were similar to each other and somewhat close to the bamboo soil sample, but different from other samples (**Figures 1C,D** and **Supplementary Tables S9,S10**).

Profiling of GH Families and Cellulases in the APMP Consortium
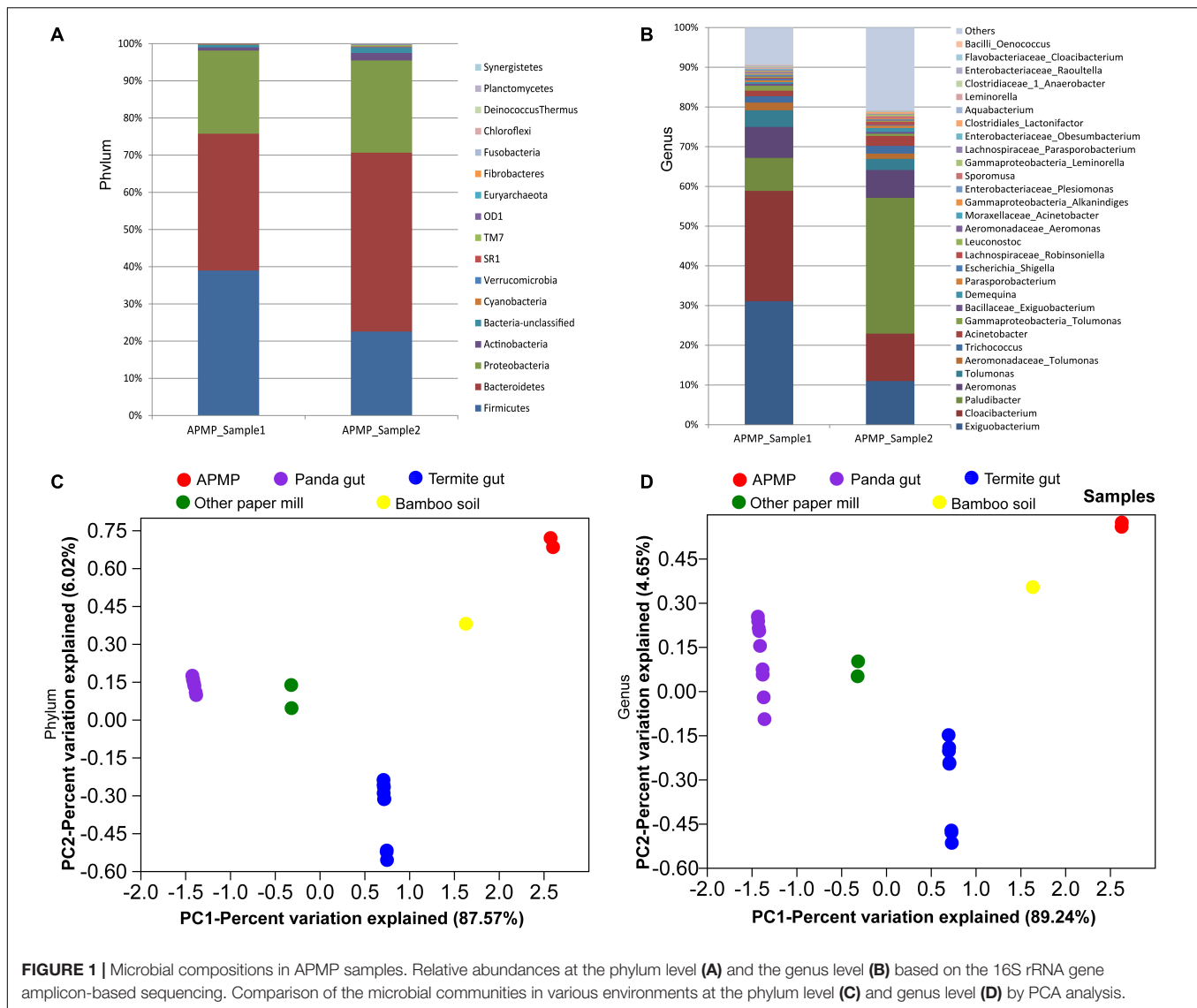
We next used a metagenomic approach based on next-generation *de novo* sequencing to identify the functional

---

[4]http://csbl.bmb.uga.edu/dbCAN
[5]www.rcsb.org
[6]http://www.tcdb.org/download.php

[7]http://amphoranet.pitgroup.org

FIGURE 1 | Microbial compositions in APMP samples. Relative abundances at the phylum level **(A)** and the genus level **(B)** based on the 16S rRNA gene amplicon-based sequencing. Comparison of the microbial communities in various environments at the phylum level **(C)** and genus level **(D)** by PCA analysis.

attributes encoded in the APMP consortium, with a focus on the cellulolytic microbes and their functional genes. The metagenomic sequencing generated 20 million and 23 million high-quality reads in samples 1 and 2, respectively. Assembly of the reads resulted in 172,009 and 186,534 total contigs, respectively (≥300 bp); among them, 23,086 contigs in sample 1 and 25,201 contigs in sample 2 are larger than 1000 bp. The largest contig is 122,944 bp in length. In total, 223,510 and 249,163 coding genes were predicted in samples 1 and 2, respectively; 55.56% of the genes in sample 1 and 55.48% of the genes in sample 2 were assigned putative functions based on the COG database. Sequencing and assembly statistics are given in **Supplementary Table S4**.

Using the dbCAN database as a reference, we assessed the functional capacities of the APMP samples and identified 5808 genes of 92 GH families in APMP sample 1, and 5868 genes of 93 GH families in APMP sample 2. In total, 98 GH families were present in APMP samples, and 87 GH families were present

in both samples. By contrast, only 448 genes of 44 GH families were identified in the giant panda gut samples (Zhu et al., 2011), and 2779 genes of 75 GH families were identified in the termite hindgut sample (Warnecke et al., 2007; **Supplementary Table S6**). The distribution of the GH families among the three types of samples is summarized in **Figure 2**. Thirty-nine GH families were present in all three samples. The majority of the identified GH families in the panda gut (42/44) and termite hindgut (72/75), respectively, were shared with the APMP samples. Twenty-three GH families were present in the APMP samples only (**Supplementary Table S6**).

Cellulases, a heterogeneous group of enzymes that mediate cellulose degradation, have been found within 13 GH families (GH 5–9, 12, 44, 45, 48, 51, 61, 74, and 124) (Sukharnikov et al., 2011; Okano et al., 2015). Notably, these families are unrelated and diverse in sequence, and their domain linkage between the GH domains and the CBM domains had been proposed (Sukharnikov et al., 2011). The architectures of the cellulase genes

FIGURE 2 | Venn diagram of the GH families in the samples of APMP, panda gut, and termite hindgut.
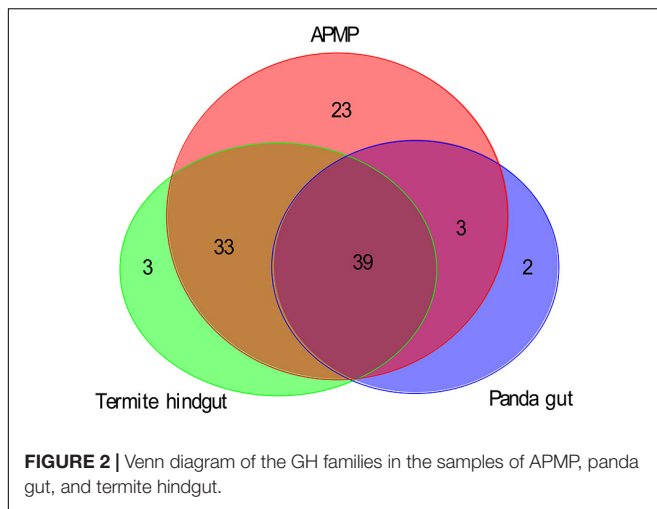
were analyzed and representative results were shown in **Figure 3**. Seven distinct CBM types (CBM 6, X2, 36, 11, 17/18, 30) were located on these genes. Notably, CBM30 was frequently linked to GH9. These cellulases have low identities (28–74%) with their counterparts from various bacterial phylotypes in the NCBI-nr database, suggesting that the APMP cellulase genes originated from divergent bacterial hosts.

All full-length genes belonging to the 13 GH families in samples 1 and 2 are listed in **Supplementary Table S7** with their homologous templates and putative functions. Two representatives, gene-40972 (GH5) and gene-6561 (GH9) showed high homologies with their homologs in the RCSB PDB database, therefore they were chosen as examples for analysis using the Galaxy template-based modeling program (Shin et al., 2014).

The cellulase encoded by gene-40972 was predicted to fold into a cylindrical $(\alpha/\beta)_8$-barrel with a catalytic pocket on the C-terminal side of the barrel, typical of GH5 cellulases. The cellulase encoded by gene-6561 was predicted to contain a catalytic domain with a GH9 signature $(\alpha/\alpha)_6$-barrel fold, and an N-terminal immunoglobulin-like (Ig-like) domain. The substrate binding pockets and the catalytic residues of gene-40972 (**Figure 4A**) and gene-6561 (**Figure 4B**) were highly similar to their structural homologs, *C. thermocellum* CtCelC and LC-CelG, respectively (Dominguez et al., 1995; Kerepesi et al., 2014; Laczny et al., 2015; Liu et al., 2015; Okano et al., 2015). In both cases, a cellotriose molecule was predicted to bind to the substrate pocket, and the catalytic residues D149, D152, and E516 of gene-6561, and E146, H204, and E287 of gene-40972 were all conserved. The homologous structures suggest they have conserved the cellulase functions.

Analysis of amino acid compositions, charges, and isoelectric points of the two genes and their homologs indicates that the cellulase of gene-40972 has similar properties to alkaline cellulase 103 from *Bacillus sp.* found in soda lakes, e.g., isoelectric points closer to 4 and negative charges lower than −20, but was different from CtCelC and the FnCel5A (**Supplementary Table S11**; Dominguez et al., 1995; Shaw et al., 2002; Zheng et al., 2012). Notably, both gene-40972 and 103 cellulases have a marked
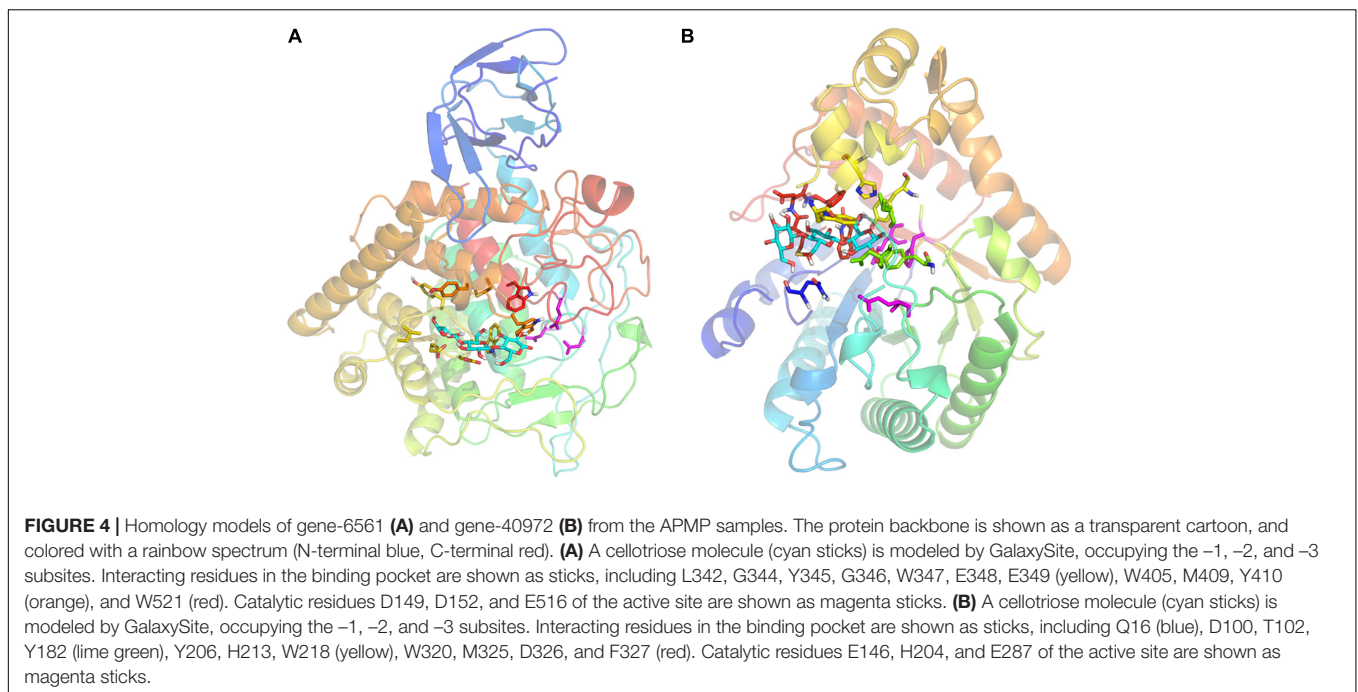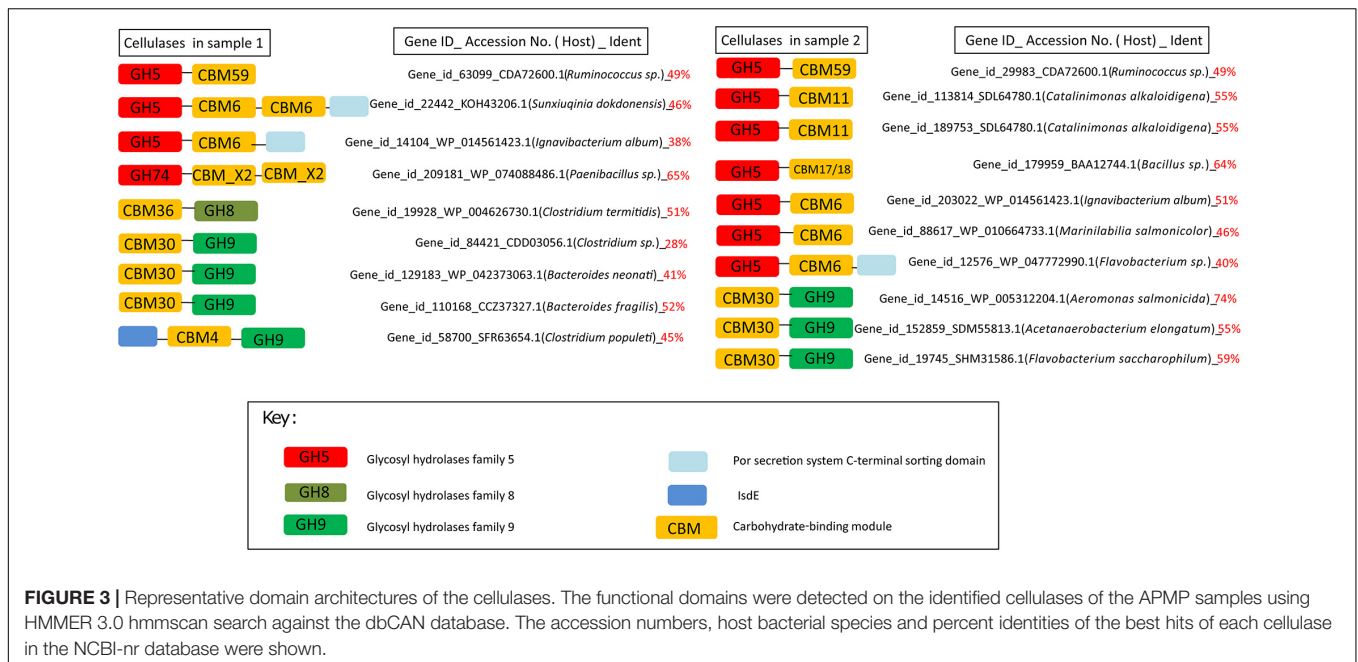
decrease in Lys residues, as was suggested to be an alkaline adaptation mechanism for the industrial cellulase K (Shirai et al., 2001). The cellulase of gene-6561 was also predicted to have an isoelectric point close to 4 and a negative charge lower than −20 (**Supplementary Table S11**). However, when compared to its structural homologs, LC-CelG, AaCel9A, and CtCelD (Chauvaux et al., 1995; Dominguez et al., 1995; Shirai et al., 2001; Shaw et al., 2002; Eckert et al., 2009; Zheng et al., 2012; Kerepesi et al., 2014; Laczny et al., 2015; Liu et al., 2015; Okano et al., 2015), gene-6561 cellulase unexpectedly has an increased number of Lys residues, and strikingly doubled the amount of Glu residues, resulting in its low isoelectric point. It is likely gene-6561 cellulase of the GH9 family employs a different alkaline adaptation mechanism from the GH5 family cellulases.

The Classification of "Beneficiaries" and "Helpers" Regarding Cellulose Degradation in the APMP Consortium

In the APMP consortium where cellulose is the dominant carbon source, microbes produce cellulases to sustain their growth. We also investigated the gene content of the metagenomes and assigned the genes taxonomic origins according to the best hits in the LBPSDB. The genera of *Cloacibacterium, Paludibacter, Exiguobacterium, Acetivibrio, Tolumonas*, and *Clostridium* have the highest numbers of protein sequences in the APMP samples (**Supplementary Table S5**). Most (95.1%) of the protein sequences encoded by the genera showed less than 90% identities with their counterparts in the LBPSDB, indicating that the APMP strains belong to novel species of the genera.

Further, the protein sequences assigned to the genera were predicted in KEGG and the proteins involved in cellulose-degrading pathways were determined. In total, 1744 carbohydrate transporter genes were identified in the APMP samples (**Supplementary Table S12**); among them, 157 genes encode cellobiose-specific PTS family homologs of CelC, CelB, and CelA (Parker and Hall, 1990), and 85 genes encode cellobiose-specific ABC family homologs of CebE, CebF, CebG, and MsiK (Schlosser et al., 1999). Accordingly, in *Clostridium*, genes encoding endoglucanase [EC:3.2.1.4], beta-glucosidase [EC:3.2.1.21], 6-phospho-beta-glucosidase [EC:3.2.1.86], as well as PTS transporter homologs of CelC, CelB, and CelA were identified, implying that *Clostridium* species likely function as helpers and can utilize cellulose by themselves. In *Paludibacter*, genes encoding endoglucanase, beta-glucosidase, and cellobiose phosphorylase were identified, but they lacked related transporters. Note that only two bacterial cellodextrin transporter families have been characterized and reported to date (Parker and Hall, 1990; Schlosser et al., 1999); thus, it is likely that *Paludibacter* species encode an unclassified cellodextrin transporter and also function as helpers.

In *Cloacibacterium, Exiguobacterium, Acetivibrio*, and *Tolumonas*, genes encoding intracellular-process-related enzymes such as beta-glucosidases and 6-phospho-beta-glucosidases were identified, but endoglucanases were absent, suggesting that these four genera function as beneficiaries in the APMP consortium. The genes encoding cellodextrin transporter
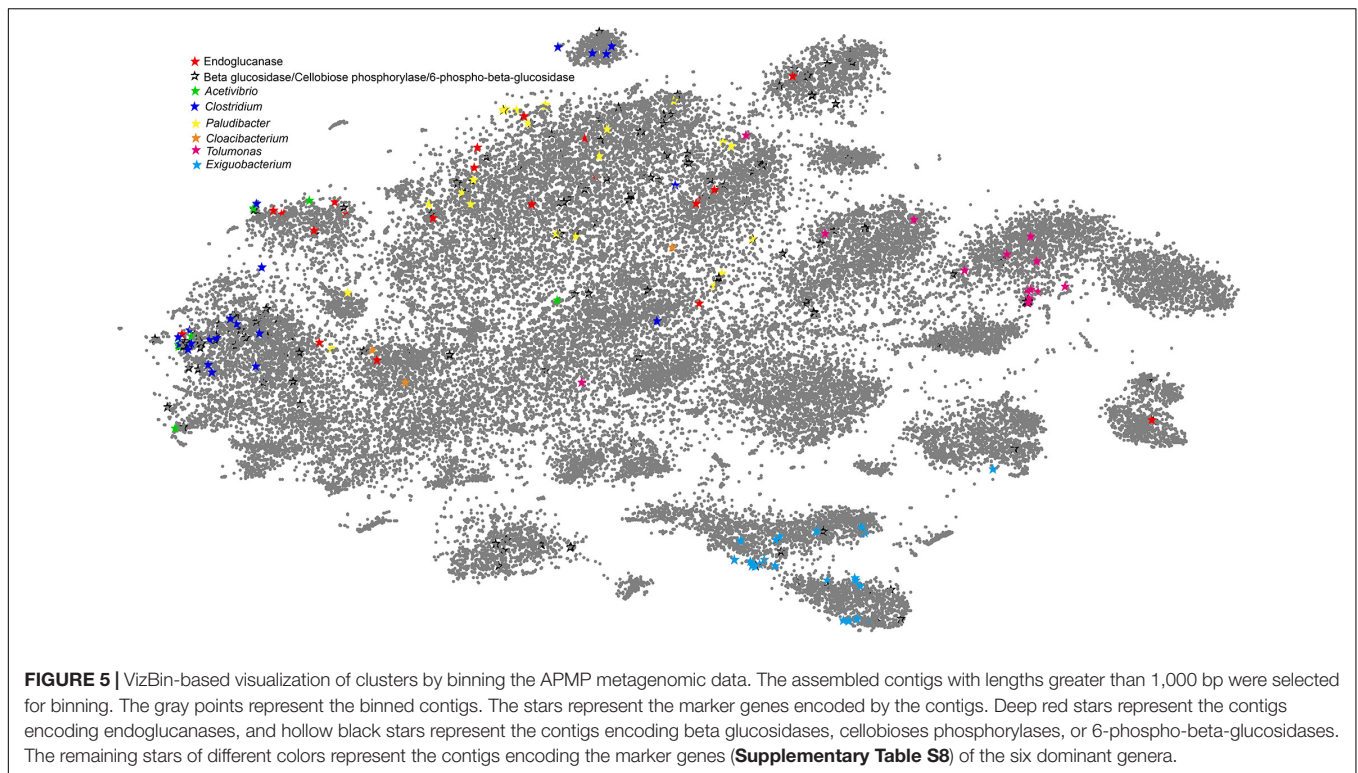
FIGURE 3 | Representative domain architectures of the cellulases. The functional domains were detected on the identified cellulases of the APMP samples using HMMER 3.0 hmmscan search against the dbCAN database. The accession numbers, host bacterial species and percent identities of the best hits of each cellulase in the NCBI-nr database were shown.

FIGURE 4 | Homology models of gene-6561 **(A)** and gene-40972 **(B)** from the APMP samples. The protein backbone is shown as a transparent cartoon, and colored with a rainbow spectrum (N-terminal blue, C-terminal red). **(A)** A cellotriose molecule (cyan sticks) is modeled by GalaxySite, occupying the −1, −2, and −3 subsites. Interacting residues in the binding pocket are shown as sticks, including L342, G344, Y345, G346, W347, E348, E349 (yellow), W405, M409, Y410 (orange), and W521 (red). Catalytic residues D149, D152, and E516 of the active site are shown as magenta sticks. **(B)** A cellotriose molecule (cyan sticks) is modeled by GalaxySite, occupying the −1, −2, and −3 subsites. Interacting residues in the binding pocket are shown as sticks, including Q16 (blue), D100, T102, Y182 (lime green), Y206, H213, W218 (yellow), W320, M325, D326, and F327 (red). Catalytic residues E146, H204, and E287 of the active site are shown as magenta sticks.

systems, such as homologs of CelC, CelB, and CelA, were also found in *Exiguobacterium, Acetivibrio*, and *Tolumonas*, but were absent in *Cloacibacterium* possibly due to low homology.

Thus, our analysis suggests that these dominant genera play distinct roles in cellulose-degrading process in APMP consortium. The genera *Clostridium and Paludibacter* can produce cellulases generating cellodextrin products that diffuse and are shared by the whole community. The other four genera survive by importing and assimilating the generated cellodextrin products. According to the BQH, the genera *Clostridium* and

Paludibacter can be designated as "helpers," the other four genera as "beneficiaries" (Morris et al., 2012).

Binning of the Metagenomic Data Supporting the Classification of Beneficiaries and Helpers

Similarity-based taxonomic assignment of the genes predicted from the environmental metagenomic dataset may result in bias because of a limited number of bacteria with sequenced genomes

FIGURE 5 | VizBin-based visualization of clusters by binning the APMP metagenomic data. The assembled contigs with lengths greater than 1,000 bp were selected for binning. The gray points represent the binned contigs. The stars represent the marker genes encoded by the contigs. Deep red stars represent the contigs encoding endoglucanases, and hollow black stars represent the contigs encoding beta glucosidases, cellobioses phosphorylases, or 6-phospho-beta-glucosidases. The remaining stars of different colors represent the contigs encoding the marker genes (**Supplementary Table S8**) of the six dominant genera.

so far. To validate the above observations, we combined the sequencing data of the two APMP samples and employed a reference-independent binning approach to cluster the assembled microbial metagenomic contigs. The contigs were labeled using the phylogenetic marker genes (**Supplementary Table S8**) and the genes corresponding to cellulose-degradation derived from the APMP metagenomic dataset (**Figure 5**).

In **Figure 5**, the clusters affiliated to *Clostridium, Paludibacter, Exiguobacterium,* and *Tolumonas*, respectively, can be readily recognized as we can see that the contigs with the corresponding phylogenetic marker genes intensively appear in distinct clusters. *Paludibacter* is one of the most abundant genera. In **Figure 5**, the marker genes of the *Paludibacter* genus lie in the largest cluster, but highly scattered distribution of the marker genes likely suggests that *Paludibacter* species are diverse in the consortium. This case also can be observed that the phylogenetic marker genes of *Cloacibacterium* and *Acetivibrio* scatter among different clusters. Significantly, most of the cellulose-degrading genes are adjacent to the marker genes of the six genera in **Figure 5**, suggesting that the cellulose-degrading genes are mainly carried by these top abundant genera, the main players in the APMP niche.
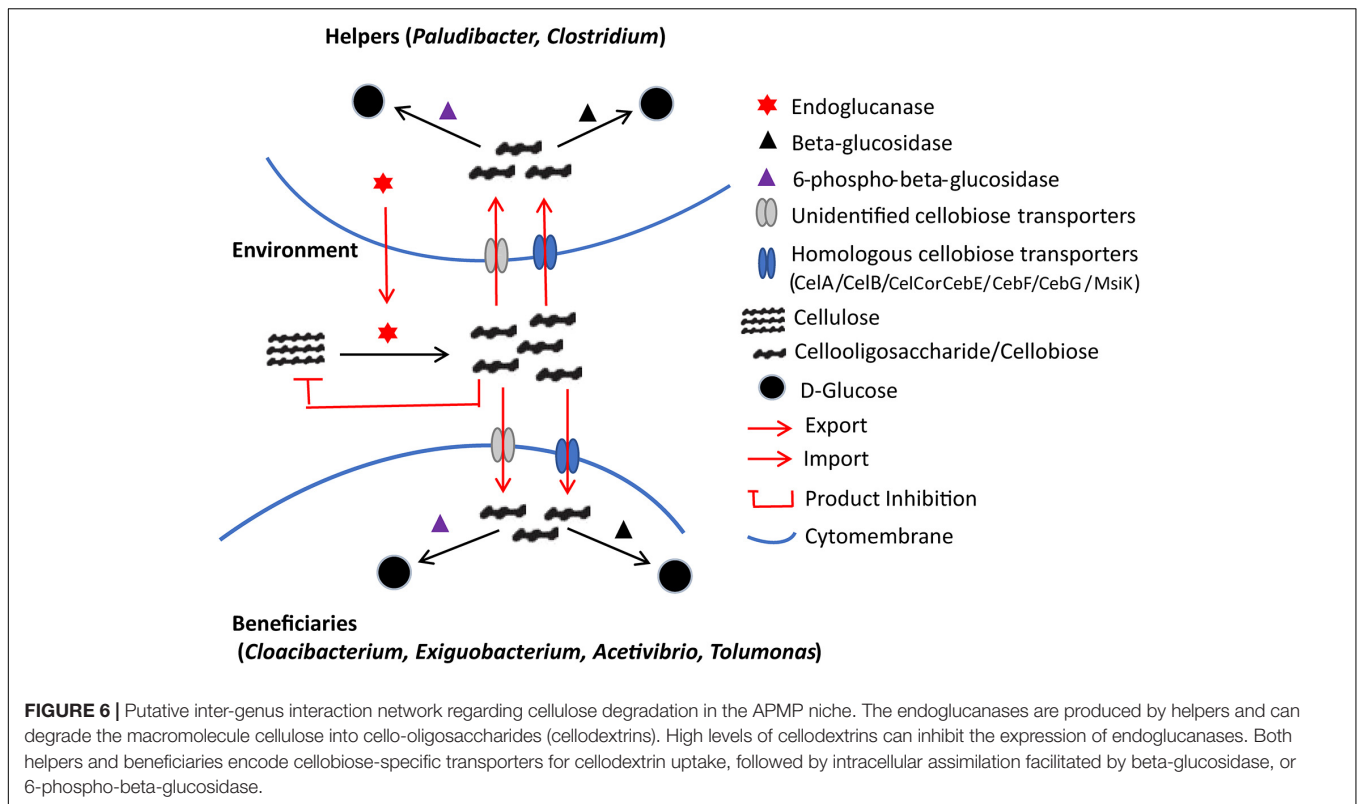
Most of the marker genes affiliated with *Clostridium* are distributed in two distinct clusters, implying that the contigs belong to at least two different *Clostridium* groups. Interestingly, only one cluster contains the genes of endoglucanases and downstream enzymes for cellulolysis, indicating that this species functions as a helper and the other group does not. Both upstream and downstream cellulose-degrading enzyme genes are also shown in the clusters of *Paludibacter*, verifying

that *Clostridium* and *Paludibacter* play a role as helper in the process of cellulose degradation in the consortium. The marker genes affiliated with *Exiguobacterium* and *Tolumonas* lie in the clusters that lack endoglucanase contigs but encode various intracellular-process-related enzymes such as beta-glucosidases, cellobiose phosphorylase, and 6-phospho-beta-glucosidases, strongly supporting our classification of *Exiguobacterium* and *Tolumonas* as beneficiaries.

DISCUSSION

In the ancient Chinese paper-making procedure, bamboo fibers are softened and soaked in lime water, which would have efficiently extracted the crude hemicellulose (Peng and She, 2014), but retained the crystalline cellulose. A cellulose-rich and alkaline niche has been generated unintentionally in this process, and arguably a microbial community including numerous bacteria and fungi has been evolving intermittently for 7 centuries in the isolated habitat. In this study, we investigated the microbial consortium in such an environment to understand how the microbes interplay *in situ* in cellulose-degradation process using high throughput sequencing. Due to the large size of the fungal genomes, the great diversity of environmental fungi and a small number of fungal reference genomes available, whole genomic analysis of the fungal component in a given microbiome still remains challenging in nowadays. Therefore, considering the complexity of the APMP metagenomic data, we set up a series of stringent criteria in the bioinformatic analyses to focus on bacterial component only and exclude

FIGURE 6 | Putative inter-genus interaction network regarding cellulose degradation in the APMP niche. The endoglucanases are produced by helpers and can degrade the macromolecule cellulose into cello-oligosaccharides (cellodextrins). High levels of cellodextrins can inhibit the expression of endoglucanases. Both helpers and beneficiaries encode cellobiose-specific transporters for cellodextrin uptake, followed by intracellular assimilation facilitated by beta-glucosidase, or 6-phospho-beta-glucosidase.

possible errors or bias caused by sequencing depth, wrong assembly, etc. By recruiting 16S rRNA gene reads from the APMP metagenomic sequencing data and assigning the 16S rRNA gene reads taxonomically, we estimated that the reads belonging to the top six abundant genera, including *Cloacibacterium, Paludibacter, Exiguobacterium, Acetivibrio, Tolumonas*, and *Clostridium*, occupy more than 90% of all 16S rRNA gene reads (data not shown), suggesting that the sequencing data may reach more than 100 x depth coverage of the genomes of the top abundant bacteria. The high depth coverage also can be seen from **Supplementary Table S5** that the top abundant genera have high number of hits in APMP metagenomic data. To our knowledge, this study firstly reveals that a unique microbial community containing six predominant bacterial genera is present in this isolated habitat and we provide numerous genetic evidences illustrating that these bacteria functionally cooperate with each other in the cellulose-degradation process.

The microbes of this unique niche are presumably novel, because most of the amplicon-based partial 16S rRNA gene sequencing reads were aligned with low identity (less than 90%) against the 16S rRNA gene sequences of the latest RDP Release data[8] (data not shown). The well-developed software or algorithm, such as QIIME, MetaPhlAn, and so on, may not be suitable for our sequencing data (Kuczynski et al., 2012; Segata et al., 2012). Thus, we developed in-house bioinformatic pipelines for our sequencing data. Firstly, we performed amplicon-based 16S rRNA gene sequencing and developed a bioinformatic

pipeline using Naïve Bayesian Classifier to profile the microbial composition on genus-level and phylum-level, respectively, resulting in identifying a unique microbial consortium in the APMP samples compared to other cellulose-rich samples (Demain et al., 2005; Warnecke et al., 2007; Carroll and Somerville, 2009; Zhu et al., 2011; Morris et al., 2012; He et al., 2013; Hummert et al., 2014; Lu and Lu, 2014; Zhou et al., 2014; Dash et al., 2016). Five top abundant genera in the APMP samples were *Cloacibacterium, Paludibacter, Exiguobacterium, Tolumonas*, and *Aeromonas* (**Figure 1**). Secondly, we performed metagenomic sequencing of the APMP samples to identify the gene content and bacterial taxa, demonstrating that the genera *Cloacibacterium, Paludibacter, Exiguobacterium, Acetivibrio, Tolumonas*, and *Clostridium* are dominant in the APMP samples. Among 672 genera in the database LBPSDB, each of these 6 genera has large number of genes with hits to the genes encoded by the large contigs ($\geq$300 bp) in the APMP metagenomic data (**Supplementary Table S5**). This result is not completely consistent with that of 16S rRNA gene sequencing data in that none of the abundant taxa in **Figure 1** are designated as *Clostridium* or as *Acetivibrio*. This discrepancy can be explained by the limitations of bioinformatic algorithms in identifying novel bacterial taxa from a complex microbial consortium. The key software RDP naïve Bayesian Classifier involved in our bioinformatic pipeline for analyzing 16S rRNA gene sequencing data is trained on the known type strains of 16S rRNA gene sequences (Cole et al., 2005). In contrast, the database LPBSDB that we constructed in this study contains all available 7806 whole genome sequences in NCBI, and these genomes originate from

---

[8]https://rdp.cme.msu.edu

672 genera. The genera *Clostridium* and *Acetivibrio* that were not shown in **Figure 1** have 36 and 2 genomes in the database LPBSDB, respectively. The genes of *Clostridium* and *Acetivibrio* have large numbers of best hits (28742 and 2960, respectively) to the genes encoded by the APMP metagenomic contigs. However, the mean BLAST identities for hits designated as *Clostridium* (around 64%) and *Acetivibrio* (around 61%) are considerably lower than mean identites for the other four dominant genus-level taxa identified by this method (70–86%; **Supplementary Table S5**), suggesting that those bacteria in APMP samples are likely novel but most closely related to the known strains of the genera *Clostidium* and *Acetivibrio* among the 672 genera in the database of LBPSDB, respectively.

Based on the results of taxonomic assignment and functional annotation of the genes encoded by the APMP contigs (≥300 bp), *Paludibacter* and *Clostridium* species are designated as helpers that encode the entire pathway enzymes to break down cellulose. In contrast, *Cloacibacterium, Exiguobacterium, Acetivibrio,* and *Tolumonas* lack the genes of endoglucanses and are designated as beneficiaries who benefit from the public cellodextrins produced by *Paludibacter* and *Clostridium* species. This finding was further validated by the binning analysis using the software Vizbin (**Figure 5**). Note that the contigs larger than 1000 bp account for 13% of the contigs larger than 300 bp in the two samples. In the binning analysis, only the contigs larger than 1000 bp were involved, resulting in that a small number of the marker genes are shown in **Figure 5**. Even so, the binning result significantly verifies that in the microbial consortium, the cellulose-degrading genes are mainly carried by the strains belonging to the six top abundant genera and strongly supports our classification of *Exiguobacterium* and *Tolumonas* as beneficiaries, and *Paludibacter* and *Clostridium* as helpers. However, ambiguous clusters for the contigs of *Acetivibrio* and *Cloacibacterium* shown in **Figure 5** indicate the limitations of the binning method (Parker and Hall, 1990; Chauvaux et al., 1995; Dominguez et al., 1995; Schlosser et al., 1999; Shirai et al., 2001; Shaw et al., 2002; Cole et al., 2005; Droge and McHardy, 2012; Kuczynski et al., 2012; Segata et al., 2012; Zheng et al., 2012; Kerepesi et al., 2014; Peng and She, 2014; Laczny et al., 2015).

Thus, according to the results of the metagenomic analyses, a putative inter-genera interaction network is shown in **Figure 6**. The evolution of functional dependency of beneficiaries on helpers generates commensalistic or even mutualistic interactions. Notably, cellobiose products are known to inhibit the cellulolytic process due to the product inhibition effect at moderate to high concentrations (Voutilainen et al., 2008). Therefore, their fast removal by the beneficiaries should alleviate the inhibition (Galazka et al., 2010). The beneficiaries and the helpers likely both get benefits in the APMP consortium, and their interactions can be considered not merely commensalistic but also mutualistic (Schuster et al., 2010).

CONCLUSION

Cellulose is a sustainable, cost-effective source of carbon and energy (Carroll and Somerville, 2009). Aside from its use in paper production, cellulose is also under investigation for industrial use, such as conversion into biofuels. The cellulases identified in the APMP consortium may be applied to alkaline pretreatments of cellulosic materials (Wyman et al., 2005), and the new cellulolytic species and the helper/beneficiary strategy revealed by this study may be applied to improve the future design of synthetic microbial consortia, to facilitate engineering of resilient communities, and to achieve high efficiency for industrial cellulose conversion.

DATA AVAILABILITY

All data are fully available without restriction. All the sequencing data are available from the NCBI SRA database with accession numbers SAMN07807411, SAMN07807412, and SAMN08818942. All data generated or analyzed during this study are included in this published article and its additional files.

ETHICS STATEMENT

All the samples were obtained with consent of the owner of the paper-making plant for scientific research use only and therefore no specific ethical approval is needed.

AUTHOR CONTRIBUTIONS

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2019.00618/full#supplementary-material

REFERENCES

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The carbohydrate-active enzymes database (cazy): an expert resource for glycogenomics. *Nucleic Acids Res.* 37, D233–D238. doi: 10.1093/nar/gkn663

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303

Carroll, A., and Somerville, C. (2009). Cellulosic biofuels. *Annu. Rev. Plant Biol.* 60, 165–182. doi: 10.1146/annurev.arplant.043008.092125

Chauvaux, S., Souchon, H., Alzari, P. M., Chariot, P., and Beguin, P. (1995). Structural and functional-analysis of the metal-binding sites of clostridium-thermocellum endoglucanase celd. *J. Biol. Chem.* 270, 9757–9762. doi: 10.1074/jbc.270.17.9757

Cole, J. R., Chai, B., Farris, R. J., Wang, Q., Kulam, S. A., McGarrell, D. M., et al. (2005). The ribosomal database project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Res.* 33, D294–D296.

Dash, S., Ng, C. Y., and Maranas, C. D. (2016). Metabolic modeling of clostridia: current developments and applications. *FEMS Microbiol. Lett.* 363:fnw004. doi: 10.1093/femsle/fnw004

Demain, A. L., Newcomb, M., and Wu, J. H. (2005). Cellulase, clostridia, and ethanol. *Microbiol. Mol. Biol. Rev.*69, 124–154. doi: 10.1128/MMBR.69.1.124-154.2005

Dominguez, R., Souchon, H., Lascombe, M., and Alzari, P. M. (1996). The crystal structure of a family 5 endoglucanase mutant in complexed and uncomplexed forms reveals an induced fit activation mechanism. *J. Mol. Biol.* 257, 1042–1051. doi: 10.1006/jmbi.1996.0222

Dominguez, R., Souchon, H., Spinelli, S., Dauter, Z., Wilson, K. S., Chauvaux, S., et al. (1995). A common protein fold and similar active site in two distinct families of beta-glycanases. *Nat. Struct. Biol.* 2, 569–576. doi: 10.1038/nsb0795-569

Droge, J., and McHardy, A. C. (2012). Taxonomic binning of metagenome samples generated by next-generation sequencing technologies. *Brief. Bioinform.* 13, 646–655. doi: 10.1093/bib/bbs031

Eckert, K., Vigouroux, A., Lo Leggio, L., and Morera, S. (2009). Crystal structures of A. *acidocaldarius endoglucanase* Cel9A in complex with cello-oligosaccharides: strong -1 and -2 subsites mimic cellobiohydrolase activity. *J. Mol. Biol.* 394, 61–70. doi: 10.1016/j.jmb.2009.08.060

Galazka, J. M., Tian, C., Beeson, W. T., Martinez, B., Glass, N. L., and Cate, J. H. (2010). Cellodextrin transport in yeast for improved biofuel production. *Science* 330, 84–86. doi: 10.1126/science.1192838

Galperin, M. Y., Makarova, K. S., Wolf, Y. I., and Koonin, E. V. (2015). Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.* 43, D261–D269. doi: 10.1093/nar/gku1223

Gilcreas, F. W. (1967). Future of standard methods for the examination of water and wastewater. *Health Lab. Sci.* 4, 137–141.

Hammer, O. (2001). PAST : paleontological statistics software package for education and data analysis. *Palaeontol. Electron.* 4, 1–9.

He, S., Ivanova, N., Kirton, E., Allgaier, M., Bergin, C., Scheffrahn, R. H., et al. (2013). Comparative metagenomic and metatranscriptomic analysis of hindgut paunch microbiota in wood- and dung-feeding higher termites. *PLoS One* 8:e61126. doi: 10.1371/journal.pone.0061126

Hummert, S., Bohl, K., Basanta, D., Deutsch, A., Werner, S., Theissen, G., et al. (2014). Evolutionary game theory: cells as players. *Mol. Biosyst.* 10, 3044–3065. doi: 10.1039/c3mb70602h

Kanehisa, M., Sato, Y., and Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* 428, 726–731. doi: 10.1016/j.jmb.2015.11.006

Kato, S., Haruta, S., Cui, Z. J., Ishii, M., and Igarashi, Y. (2004). Effective cellulose degradation by a mixed-culture system composed of a cellulolytic *Clostridium* and aerobic non-cellulolytic bacteria. *FEMS Microbiol. Ecol.* 51, 133–142. doi: 10.1016/j.femsec.2004.07.015

Kato, S., Haruta, S., Cui, Z. J., Ishii, M., and Igarashi, Y. (2005). Stable coexistence of five bacterial strains as a cellulose-degrading community. *Appl. Environ. Microbiol.* 71, 7099–7106. doi: 10.1128/AEM.71.11.7099-7106.2005

Kerepesi, C., Banky, D., and Grolmusz, V. (2014). AmphoraNet: the webserver implementation of the AMPHORA2 metagenomic workflow suite. *Gene* 533, 538–540. doi: 10.1016/j.gene.2013.10.015

Klemm, D., Heublein, B., Fink, H. P., and Bohn, A. (2005). Cellulose: fascinating biopolymer and sustainable raw material. *Angew. Chem.* 44, 3358–3393. doi: 10.1002/anie.200460587

Kuczynski, J., Stombaugh, J., Walters, W. A., Gonzalez, A., Caporaso, J. G., and Knight, R. (2012). Using QIIME to analyze 16S rRNA gene sequences from microbial communities. *Curr. Protoc. Microbiol.* 27, 1E.5.1–1E.5.20. doi: 10.1002/9780471729259.mc01e05s27

Kudo, H., Cheng, K. J., and Costerton, J. W. (1987). Interactions between *Treponema bryantii* and cellulolytic bacteria in the in vitro degradation of straw cellulose. *Can. J. Microbiol.* 33, 244–248. doi: 10.1139/m87-041

Laczny, C. C., Sternal, T., Plugaru, V., Gawron, P., Atashpendar, A., Margossian, H. H., et al. (2015). VizBin - an application for reference-independent visualization and human-augmented binning of metagenomic data. *Microbiome* 3, 1. doi: 10.1186/s40168-014-0066-1

Li, W., Cowley, A., Uludag, M., Gur, T., McWilliam, H., Squizzato, S., et al. (2015). The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res.* 43, W580–W584. doi: 10.1093/nar/gkv279

Lin, Y. T., Whitman, W. B., Coleman, D. C., Shi, S. Y., Tang, S. L., and Chiu, C. Y. (2015). Changes of soil bacterial communities in bamboo plantations at different elevations. *FEMS Microbiol. Ecol.* 91:fiv033. doi: 10.1093/femsec/fiv033

Liu, W., Xie, Y., Ma, J., Luo, X., Nie, P., Zuo, Z., et al. (2015). IBS: an illustrator for the presentation and visualization of biological sequences. *Bioinformatics* 31, 3359–3361. doi: 10.1093/bioinformatics/btv362

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014a). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, D490–D495. doi: 10.1093/nar/gkt1178

Lombard, V., Ramulu, H. G., Drula, E., Coutinho, P. M., and Henrissat, B. (2014b). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, D490–D495. doi: 10.1093/nar/gkt1178

Lu, X. M., and Lu, P. Z. (2014). Effects of farmhouse hotel and paper mill effluents on bacterial community structures in sediment and surface water of Nanxi River, China. *Microb. Ecol.* 68, 773–784. doi: 10.1007/s00248-014-0456-8

Morris, J. J., Lenski, R. E., and Zinser, E. R. (2012). The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. *mBio* 3:e00036-12. doi: 10.1128/mBio.00036-12

Namiki, T., Hachiya, T., Tanaka, H., and Sakakibara, Y. (2012). MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Res.* 40:e155. doi: 10.1093/nar/gks678

Okano, H., Kanaya, E., Ozaki, M., Angkawidjaja, C., and Kanaya, S. (2015). Structure, activity, and stability of metagenome-derived glycoside hydrolase family 9 endoglucanase with an N-terminal Ig-like domain. *Protein sci.* 24, 408–419. doi: 10.1002/pro.2632

Parker, L. L., and Hall, B. G. (1990). Characterization and nucleotide-sequence of the cryptic cel operon of *Escherichia-coli* K12. *Genetics* 124, 455–471.

Peng, P., and She, D. (2014). Isolation, structural characterization, and potential applications of hemicelluloses from bamboo: a review. *Carbohydr. Polym.* 112, 701–720. doi: 10.1016/j.carbpol.2014.06.068

Ravel, J., Gajer, P., Abdo, Z., Schneider, G. M., Koenig, S. S., McCulle, S. L., et al. (2011). Vaginal microbiome of reproductive-age women. *Proc. Natl. Acad. Sci. U.S.A.* 108, 4680–4687. doi: 10.1073/pnas.1002611107

Salimi, F., Zhuang, K., and Mahadevan, R. (2010). Genome-scale metabolic modeling of a clostridial co-culture for consolidated bioprocessing. *Biotechnol. J.* 5, 726–738. doi: 10.1002/biot.201000159

Schlosser, A., Jantos, J., Hackmann, K., and Schrempf, H. (1999). Characterization of the binding protein-dependent cellobiose and cellotriose transport system of the cellulose degrader *Streptomyces reticuli*. *Appl. Environ. Microb.* 65, 2636–2643.

Schuster, S., Kreft, J. U., Brenner, N., Wessely, F., Theissen, G., Ruppins, E., et al. (2010). Cooperation and cheating in microbial exoenzyme production - theoretical analysis for biotechnological applications. *Biotechnol. J.* 5, 751–758. doi: 10.1002/biot.200900303

Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., and Huttenhower, C. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat. Methods* 9, 811–814. doi: 10.1038/nmeth.2066

Shaw, A., Bott, R., Vonrhein, C., Bricogne, G., Power, S., and Day, A. G. (2002). A novel combination of two classic catalytic schemes. *J. Mol. Biol.* 320, 303–309. doi: 10.1016/S0022-2836(02)00387-X

Shin, W. H., Lee, G. R., Heo, L., Lee, H., and Seok, C. (2014). Prediction of protein structure and interaction by GALAXY protein modeling programs. *Biodesign* 2, 1–11.

Shirai, T., Ishida, H., Noda, J., Yamane, T., Ozaki, K., Hakamada, Y., et al. (2001). Crystal structure of alkaline cellulase K: insight into the alkaline adaptation of an industrial enzyme. *J. Mol. Biol.* 310, 1079–1087. doi: 10.1006/jmbi.2001.4835

Sukharnikov, L. O., Cantwell, B. J., Podar, M., and Zhulin, I. B. (2011). Cellulases: ambiguous nonhomologous enzymes in a genomic perspective. *Trends Biotechnol.* 29, 473–479. doi: 10.1016/j.tibtech.2011.04.008

Tang, S., Antonov, I., and Borodovsky, M. (2013). MetaGeneTack: ab initio detection of frameshifts in metagenomic sequences. *Bioinformatics* 29, 114–116. doi: 10.1093/bioinformatics/bts636

Voutilainen, S. P., Puranen, T., Siika-Aho, M., Lappalainen, A., Alapuranen, M., Kallio, J., et al. (2008). Cloning, expression, and characterization of novel thermostable family 7 cellobiohydrolases. *Biotechnol. Bioeng.* 101, 515–528. doi: 10.1002/bit.21940

Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naive bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261–5267. doi: 10.1128/AEM. 00062-07

Warnecke, F., Luginbuhl, P., Ivanova, N., Ghassemian, M., Richardson, T. H., Stege, J. T., et al. (2007). Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* 450, 560–565. doi: 10. 1038/nature06269

Wyman, C. E., Dale, B. E., Elander, R. T., Holtzapple, M., Ladisch, M. R., and Lee, Y. Y. (2005). Comparative sugar recovery data from laboratory scale application of leading pretreatment technologies to corn stover. *Bioresour. Technol.* 96, 2026–2032. doi: 10.1016/j.biortech.2005. 01.018

Zheng, B. S., Yang, W., Zhao, X. Y., Wang, Y. G., Lou, Z. Y., Rao, Z. H., et al. (2012). Crystal structure of hyperthermophilic endo-beta-1,4-glucanase implications for catalytic mechanism and thermostability. *J. Biol. Chem.* 287, 8336–8346. doi: 10.1074/jbc.M111.266346

Zhou, Y., Pope, P. B., Li, S., Wen, B., Tan, F., Cheng, S., et al. (2014). Omics-based interpretation of synergism in a soil-derived cellulose-degrading microbial community. *Sci. Rep.* 4, 5288. doi: 10.1038/srep05288

Zhu, L., Wu, Q., Dai, J., Zhang, S., and Wei, F. (2011). Evidence of cellulose metabolism by the giant panda gut microbiome. *Proc. Natl. Acad. Sci. U.S.A.* 108, 17714–17719. doi: 10.1073/pnas.101795 6108