Lignocellulose degradation in *Protaetia brevitarsis* larvae digestive tract: refining on a tightly designed microbial fermentation production line

Kui Wang
Institute of Plant Protection, Chinese Academy of Agricultural Sciences
Peiwen Gao
Institute of Plant Protection, Chinese Academy of Agricultural Sciences
Lili Geng
Institute of Plant Protection, Chinese Academy of Agricultural Sciences
Chunqin Liu
Cangzhou Academy of Agricultural and Forestry Sciences
Jie Zhang
Institute of Plant Protection, Chinese Academy of Agricultural Sciences
Changlong Shu (clshu@ippcaas.cn)
Chinese Academy of Agricultural Sciences Institute of Plant Protection https://orcid.org/0000-0002-6118-9981

Research

Keywords: Lignocellulose degradation, Protaetia brevitarsis, transcriptome, microbiome, CAZymes

Posted Date: November 8th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1049831/v1

Abstract

Background

The Scarabaeidae insect *Protaetia brevitarsis* (PB) has recently gained increasing research interest as resource insects, because its larvae can effectively use various organic matter ranging from decaying plant residues to humus and livestock waste and convert those compounds to nonphytotoxic and high humic acid content frass fertilizer, as well as healthy nutritional insect protein sources. Lignocellulose is the main component of PB larvae (PBLs) food sources. PBLs show high lignin and polysaccharide degradation efficiency; however, genome annotation shows that PBLs' carbohydrate-active enzymes (CAZymes) are not able to complete the lignocellulose degradation process. Therefore, the mechanism by which PBLs efficiently degrade lignocellulose is a worthy issue for further study.

Results

Here, we used a combined host genomic and gut metagenomic datasets to investigate the lignocellulose degradation activity of PBLs. First, a comprehensive gut gene catalog comprising gut microbial genes and host gut transcriptomic genes was established. The data showed that PBLs selectively enriched lignocellulose-degrading microbial species mainly from *Firmicutes* and *Bacteroidetes*, which are capable of producing a broad array of cellulases and hemicellulases, thus playing a major role in lignocellulosic biomass degradation. Gene annotation revealed PBLs gut microbiome encoded a full spectrum of CAZymes involving in the hydrolysis of lignocellulose (including 39,969 CAZymes), while the host only expressed 33 lignocellulose degradation-related CAZymes in the guts. The PBL hosts provide lignin pretreatment processes via their evolved strong mouthparts as well as a strong alkaline environment in the midgut, which thus complements the lack of laccases in the PBL holobiont. In addition, most of the lignocellulose degradation-related gene sequences in the PBL microbiome were novel, and most of the recovered high-quality genome bins with independent (hemi)cellulose degradation capability from the PBL gut microbiome were novel species.

Conclusions

This work shows that there is a unique teamwork between PBLs and their gut bacterial flora for efficient lignocellulose degradation. PBLs are a promising model to study lignocellulose degradation, which can provide highly abundant novel enzymes and relevant lignocellulose-degrading bacterial strains for biotechnological biomass conversion industries.

Background

With the development of human civilization, animal domestication and animal agriculture continue to play an essential role in food supply, especially converting plant biomass to proteins, ranging from dairy products to beef, poultry, fish, eggs and pork [1]. Currently, great pressures imposed by global population growth have not only shifted the livestock industry toward a larger scale but also urged people to search for novel, sustainable protein sources. Domestic insects that convert agricultural waste to edible proteins have been acknowledged as a feasible strategy due to their broader range of feed sources and higher feed conversion rates (FAO, http://www.fao.org/edible-insects/en/).

The Scarabaeidae insect *Protaetia brevitarsis* (PB) is a promising alternative [2]. First, PB larvae (PBLs) are saprophagous and can feed on a large amount of organic matter, from decaying plant residues to humus, livestock waste, and spent mushroom substrate. Based on their special feeding habits, the feed source of PBLs is the most extensive. Second, PBLs can effectively digest plant residues and accumulate proteins and lipids for larval development. Dried mature PBL present 54.16-to-67.07% protein, 9.91-to-19.38% lipid and a wide variety of micronutrients [2]. In addition, our recent works demonstrated that the PBLs could effectively convert plant residues to nonphytotoxic and high humic acid content frass fertilizer [3, 4]. These reports indicate that lignocellulosic biomass can be efficiently digested in the PBL digestive tract.

In nature, saprophagous scarab larvae, including PBLs, are important “litter transformers” and play an important role in the terrestrial carbon cycle. They have evolved a highly compartmented digestive tract to help them obtain nutrients and energy from lignocellulose. The typical alimentary tract of scarab larvae is divided into three major sections: a foregut used for food storage, a long midgut occupying most of the length of the body cavity, and a modified expanded hindgut which is often referred to as a fermentation chamber [5]. Like termites, scarab larvae possess a highly alkaline midgut, which is believed to help to increase the solubility of organic polymers, thus rendering the organic components accessible for digestion in subsequent less-alkaline compartments [5, 6]; an enlarged hindgut with a near-neutral environment possessing a dense and diverse microbial community, analogous to the microorganism-rich rumen of higher mammals, is the primary site of microbial fermentation for lignocellulose digestion [5].

Lignocellulose is mainly composed of lignin, cellulose, and hemicellulose, forming a highly complex and varying polymeric structure that is highly recalcitrant to degradation and thus requires a consortium of carbohydrate-active enzymes (CAZymes) acting in synergism for its complete deconstruction [7]. Lignin degradation is an enzymatic oxidation catalyzed by two main groups of enzymes: lignin-modifying enzymes (LMEs) and lignin-degrading auxiliary (LDA) enzymes [8]. In contrast to lignin, cellulose and hemicellulose enzymatic degradation is mainly mediated by the hydrolysis process through the action of glycoside hydrolases. The process of cellulose degradation involves synergistic actions of three types of cellulases, endoglucanase, exoglucanase and β-glucosidase, while the depolymerization of hemicellulose requires endo-hemicellulases, exo-hemicellulases, and debranching enzymes that cleave the side chains of the polymers or associated oligosaccharides [7, 9]. In addition, recent studies have indicated that the lignocellulose degradation efficiency can be remarkably improved by the cooperative action of lytic polysaccharide monooxygenases (LPMOs), which are able to directly oxidize and depolymerize insoluble crystalline substrate surfaces or soluble hemicellulosic substrates such as xyloglucan, xylan, and β-glucans [10-12]. In natural ecosystems, degradation of lignocellulosic biomass is mainly dependent on a repertoire of enzymes produced by bacteria and fungi [7, 8]; however, more efficient degradation is usually achieved at the holobiont level, i.e., relying on complementary and synergistic cooperation between the host and its microbial symbionts [13-15]. Compared with other lignocellulose decomposers, such as the termite gut [16, 17], earthworm gut [18], or cattle rumen [19], PBLs possess more abundant gut microbial communities [20], indicating abundant and novel lignocellulosic enzymes or microbial candidates in PBL gut ecosystem. In the present work, to provide new insights into the mechanisms underlying highly efficient lignocellulose degradation by PBLs, we combined genomic, transcriptomic and metagenomic approaches at the holobiont level for the first time to (□) identify the lignocellulose-degrading CAZymes and lignocellulose-binding modules present in both the host and microbiota, (□) characterize the microbial taxa contributing the lignocellulose degradation-related genes, and (□) recover the individual lignocellulolytic species in the microbiota through a metagenomic binning approach. These

investigations can not only improve our understanding of the lignocellulose degradation mechanisms in PBLs but also contribute to applications in farming edible insects as well as biofuel and biomaterial production.

Methods

Preparation of samples

The PB laboratory population was derived from a field population collected in Gongzhuling, Jilin Province, China [21], and reared in a constant environment in an incubator at 26℃, 40-60% relative humidity, and a photoperiod of 12 h light:12 h dark. The larvae were fed corn straw, which was crushed into approximately 1 cm pieces with 50% moisture content. Third-instar larvae were selected and chilled on ice for dissection. After surface sterilization using 70% ethanol, the midgut and hindgut were dissected for subsequent analysis.

PBL gut transcriptome analysis

To prepare midgut and hindgut tissues, the dissected midgut and hindgut were washed in a cold 125 mM NaCl solution after removing gut contents. Subsequently, the washed gut tissue was transferred into a homogenizer and homogenized with TRIzol reagent (Invitrogen, USA), and RNA was extracted according to the manufacturer's protocol. RNA quality and quantity were determined with gel electrophoresis and a NanoDrop spectrophotometer (Thermo Fisher, USA). Then, RNA sequencing libraries were generated using an Illumina TruSeq Stranded mRNA Library Prep Kit (Illumina, USA), and sequencing was performed on an Illumina HiSeq 2500 sequencer (Illumina, USA) to produce 2 × 150 bp paired-end reads. When the raw reads were produced, quality control, adapter trimming and quality filtering were performed by Fastp (version 0.21.0) [22]. Finally, clean reads were deposited in the NCBI Sequence Read Archive (SRA). The SRA accessions SRR5038971, SRR5039436, SRR5039445, and SRR14128221, SRR14132028, SRR14132050 correspond to midgut and hindgut samples from three larvae, respectively (Additional file 1: Table S1).

To determine the transcriptomic expression profiles, the clean reads were aligned to the PB reference genome [21] using Spliced Transcripts Alignment to a Reference (STAR) [23], and StringTIE [24] was employed to perform expression analysis. Each gene expression level was normalized to its length for each replicate using the reads per kilobase per million (RPKM) method, which eliminates the influence of varying gene lengths and sequencing discrepancies in the calculation of gene expression. T-tests were used to evaluate the significance of the differences between the RPKM values of the midgut and hindgut sample groups.

Gut metagenome sequencing and assembly

To prepare enough DNA for gut metagenomic sequencing, ten 3rd-instar PBLs fed with corn straw were dissected, and the midgut and hindgut contents were pooled together. The DNA of the pooled midgut or hindgut contents was extracted using an Axyprep Multisource Genomic DNA Miniprep Kit (AxyGen, USA).

Then, paired-end TruSeq DNA PCR-Free libraries with insert sizes of 250 and 420 bp were constructed from the samples, and an Illumina HiSeq 2500 sequencer (Illumina, USA) was used to sequence the libraries to produce 2 × 150 bp paired-end reads. The quality of raw reads was checked with FastQC (version 0.11.9, https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Then, the adapters were trimmed using Trimmomatic (version 0.39) [25]. Reads shorter than 36 bp were removed. The clean reads were deposited to the SRA, where the SRA accessions SRR14139157-SRR14179698 were midgut samples and SRR14150473-

SRR14209386 were hindgut samples (Additional file 1: Table S1). To assist genome binning, available PBL hindgut (AHG, BHG, and MHG) and frass (AFR, BFR, and MFR) metagenomes in the SRA database were also employed (Additional file 1: Table S1).

High-quality reads from midgut and hindgut samples were pooled, and the metagenomic classifier MetaPhlAn (version 3.0.13) [26] was employed for profiling all the reads in the community with default parameters to infer the taxonomic composition of the microbial community. Subsequently, MEGAHIT (version 1.2.9) [27] was used for coassembly. Following assembly, reads were mapped to assembled contigs using the BWA-MEM algorithm [28] and SAMtools (version 1.6) [29] to obtain coverage information. Contig binning was conducted to recover individual genomes based on both tetranucleotide frequencies and sequence coverages. MaxBin 2.0 (version 2.2.7) [30] and MetaBAT 2 (version 2.12.1) [31] were used for independent binning using contigs longer than 1,500 bp and clean reads from each sample. All generated bins were aggregated and then dereplicated using dREP (version 1.4.3) [32] with default parameters. CheckM (version 1.0.7) [33] was used to estimate the genome completeness and contamination of all dereplicated bins. The binned genomes were assigned to taxa following the procedure proposed by Stewart et al 2017 [34] using the MAGpy [35] program. The phylogenetic tree of bins was built based on a concatenated protein sequence alignment using PhyloPhlAn (version 3.0.60) [36] and was annotated through iTOL (version 5, https://itol.embl.de) [37]. The relative abundance of individual taxa was measured by mapping the clean reads against binned scaffolds after a normalization step based on the size of the relevant genome bins.

Gene functional annotation

Before functional annotation, gene prediction of metagenomic contigs or bins was performed. Protein-coding sequences (CDSs) from coassembled metagenomic contigs were predicted using Prodigal (version 2.6.3) with the option –p meta [38]. The CDSs, rRNAs, and tRNAs of each metagenomic bin were predicted using the Prokka (version 1.13) [39] included in the metaWRAP [40] pipeline, with default parameters.

In the present work, gene functional annotation was focused on the lignocellulose degradation process. The Carbohydrate-Active enZYmes (CAZy) modules were identified using the CAZy database [41]. dbCAN2 was employed for annotating CAZy families from the PBL genome, transcriptome, metagenome, and individual metagenomic bins through the HMMER search approach, with an E-value threshold of 0.0001 [42]. CAZy families include auxiliary activities (AAs), carbohydrate esterases (CEs), glycoside hydrolases (GHs), glycosyltransferases (GTs), polysaccharide lyases (PLs), and carbohydrate-binding modules (CBMs). Multiple CAZy families present in a single sequence were allowed. All protein sequences identified as CAZy modules were then imported into Hotpep [43] to predict their enzymatic activity and to confirm their implication in lignocellulose degradation.

To assign the microbial sources of lignocellulose degradation-related genes, all protein sequences identified as lignocellulose-degrading CAZymes and lignocellulose-binding modules were searched against the NCBI Non-Redundant (NR) protein database (ftp://ftp.ncbi.nlm.nih.gov/blast/db/, March 2021) using DIAMOND (version 0.9.24.125) [44] with an E-value cutoff of 0.0001. Then, the DIAMOND outputs were imported into MEGAN (version 6.21.5) [45] for taxonomic assignment on the basis of the lowest common ancestor (LCA) algorithm.

Results

General features of the PBL gut metagenome and transcriptome

For PBL gut metagenome sequencing, we obtained 1,643,212,636 and 1,985,361,688 high-quality clean reads from midgut and hindgut samples, respectively, encompassing 165,332,536,581 base pairs (bp) and 199,652,618,725 bp of sequences. These were assembled into 1,337,306 contigs coding for 2,184,816 genes in the midgut and 3,930,676 contigs coding for 6,438,643 genes in the hindgut (Table 1). These data indicated that the PBL hindgut metagenome had higher complexity and taxonomic richness than the PBL midgut.

To understand the lignocellulose degradation mechanism at the holobiont level, PBL gene transcripts in the gut were determined by transcriptome sequencing. After adapter trimming and quality filtering, a total of 104,357,624 and 134,887,484 high-quality clean reads, encompassing 13,044,703,000 and 20,171,214,779 bp of sequences, were generated from midgut and hindgut libraries, respectively. After alignment to the PB reference genome, a total of 8,505 genes were identified to be expressed in the midgut and hindgut (Table 1). Among these, 8,391 genes were expressed in the midgut, 8,159 genes were expressed in the hindgut, and 8,045 of them were expressed in both the midgut and hindgut (Additional file 2: Table S2).

Table 1 Summary statistics of PBL transcriptome and metagenome

Sample	Host			Microbiome	
	Genome [c]	Midgut transcriptome	Hindgut transcriptome	Midgut	Hindgut
Total Reads [a]	-	104,357,624	134,887,484	1,643,212,636	1,985,361,688
Total bases [a]	-	13,044,703,000	20,171,214,779	165,332,536,581	199,652,618,725
Contigs	-	-	-	1,337,306	3,930,676
N50	-	-	-	1,792	1,854
CDS	33,654	8,391	8,159	2,184,816	6,438,643
CAZy modules	700	149	142	40,338	125,682
LDMs [b]	182	40	40	12,845	40,616
Cellulases	2	1	1	280	1,031
Hemicellulases	73	14	14	6,516	18,781
Hemicellulases and cellulases	31	15	15	2,493	7,741
Ligninases	42	3	3	1,038	2,089
Lignocellulose-binding modules	34	7	7	2,518	10,974

[a] reads and bases after quality trimming

[b] LDMs, lignocellulose degradation-related modules

[c] PB genome data from Wang et al (2019)

Taxonomic composition of the PBL gut microbiota

To examine whether a unique lignocellulose-degrading microbial community was enriched in the PBL gut, taxonomic distribution based on the reads from midgut and hindgut metagenome samples was analyzed at the species level (Additional file 3: Table S3). Consistent with our previous analysis based on 16S rRNA gene pyrosequencing [20], the microbial composition of the PBL midgut and hindgut was similar, but the relative abundance was different. Our taxonomic profiling analysis indicated that the microbial communities of the midgut and hindgut were composed of six and ten phyla, respectively (> 0.1% abundance). The phyla *Firmicutes*, *Bacteroidetes* and *Proteobacteria* were the predominant bacteria in the hindgut, accounting for approximately 80% of the hindgut microbial communities. In contrast, *Firmicutes* and *Proteobacteria* were the most abundant phyla in the midgut, accounting for 76.31% of the midgut microbial communities found in this work. In addition to the above phyla, *Fusobacteria* (8.14%) and *Elusimicrobia* (6.25%) were also abundant in the hindgut but were not detected in the midgut microbial communities (Fig. 1A). At the family level, 26 and 23 families (> 0.1% abundance) were detected in the midgut and hindgut microbial communities, respectively (Additional file 3: Table S3). Bacillaceae from the phylum *Firmicutes* was the most abundant family in both the midgut and hindgut, accounting for 50.44% and 33.24% of the microbial communities, respectively. Bacteroidaceae from the phylum *Bacteroidetes* was abundant in the hindgut, accounting for 23.25% of the microbial communities; however, it represented only 0.01% in the midgut. These data suggested that many bacterial species were enriched in the hindgut.

Characterization of CAZy modules in the host and PBL gut microbiome

To identify CAZy modules from the PBL holobiont, the coding genes from the PB genome and PBL gut transcriptome and metagenome were screened against the CAZy database (http://www.cazy.org).
The results indicated that a total of 344 CAZy families were identified in the PBL holobiont (Additional file 4: Table S4). For the host, a total of 700 CAZy modules from 89 CAZy families were identified in the genome, and 149 CAZy modules from 58 CAZy families were confirmed to be expressed in the gut, where 149 modules were expressed in the midgut and 142 were expressed in the hindgut (RPKM > 0, Additional file 5: Table S5). Among these modules, 58 modules were expressed at a significantly higher level in the midgut than in the hindgut, while 26 were expressed at a significantly higher level in the hindgut than in the midgut ($p < 0.05$, Additional file 5: Table S5). For the PBL gut microbiome, a total of 166,020 CAZy modules from 343 CAZy families were identified, including 40,338 CAZy modules from the midgut microbiome and 125,682 CAZy modules from the hindgut microbiome, accounting for 1.76% and 1.75% of the total genes in the midgut and hindgut gene catalogs, respectively.

Among these CAZy modules, GTs catalyze the formation of glycosidic linkages to form glycosides and are essential for biological development and environmental adaptation. However, in the PBL holobiont, GT was not the largest family but instead constituted the second largest family. There were 95 different GT families, representing 83.33% of all known GT families in the CAZy database. A total of 44,610 GT modules were identified, including 44,453 modules and 157 modules in the microbiome and host genome, respectively. GT2 was the most prominent CAZy family in the PBL holobiont, with 15,745 modules representing 9.48% of microbiome CAZy modules and 8 modules representing 1.14% of host CAZy modules.

GHs, CEs, PLs and AAs catalyze the breakdown or modification of carbohydrates and glycoconjugates, which are important for the hydrolysis and utilization of lignocellulose. In the PBL holobiont, GHs were the dominant class and comprised 124 different families, representing approximately 72.51% of all known GH families in the CAZy

database. For the gut microbiome, a total of 52,405 GH modules were identified, including 12,565 modules in the midgut and 39,840 modules in the hindgut. In contrast, only 129 GH modules were identified in the PB genome, among which 52 were expressed in both the midgut and hindgut. In addition, 16 CE families with 19,206 modules in the microbiome and 102 modules in the host, 23 PL families with 3,224 modules in the microbiome and one in the host, and 10 AA families with 3,267 modules in the microbiome and 42 modules in the host were also identified.

CAZymes often display a modular structure with noncatalytic modules, i.e., CBMs, appended to the adjacent enzymatic modules. In the PBL holobiont, we identified 73 CBM families with 26,662 modules in the microbiome and 269 modules in the host. In addition to CBMs, 15,731 S-layer homology domains (SLHs) and 1,072 cellulosome-binding domains (548 cohesins and 524 dockerins) were also detected and identified as docking modules in the PBL gut microbiome. The presence of these modules suggested the potential for active cellulosome-mediated lignocellulose plant cell wall degradation in the PBL gut.

Lignocellulose degradation-related CAZy modules in the PBL holobiont

Then, we focused on the CAZymes known as lignocellulose-degrading enzymes (cellulases, hemicellulases and ligninases) and CBMs known as lignocellulose-binding modules among all identified CAZy families in the PBL holobiont (Fig. 2; Table 1; Additional file 6: Table S6). In total, 40,117 lignocellulose-degrading CAZymes were identified in the PBL holobiont, including 39,969 modules in the microbiome and 148 modules in the host (33 in the transcriptome) (Additional file 7). These modules were from 78 lignocellulose-degrading CAZy families composed of 59 GH families, ten CE families and nine AA families. Additionally, 13,526 lignocellulose-binding modules from 46 CBM families were found in the PBL holobiont, including 13,492 modules in the microbiome and 34 modules in the host (seven in the transcriptome). The hindgut microbiome contained most of the lignocellulose-degrading CAZymes and lignocellulose-binding modules, representing 74.10% (N=29,642) and 81.34% (N=10,974), respectively.

The modification and degradation of lignin has been identified as an essential step for efficient lignocellulosic biomass deconstruction [46], and lignin consumption is mainly accomplished by LMEs and LDA enzymes. LMEs are classified as laccases (EC 1.10.3.2; AA1), manganese-dependent peroxidases (EC 1.11.1.13; AA2), lignin peroxidases (EC 1.11.1.14; AA2), and versatile peroxidases (EC 1.11.1.16; AA2). Although LDA enzymes are unable to degrade lignin on their own, the lignin degradation process can be further enhanced by the action of these enzymes. In the present work, seven AA families (AA1, AA2, AA3, AA4, AA5, AA6, and AA7) involved in lignin modification and degradation were identified in the PBL holobiont, including 1,038 modules in the midgut microbiome and 2,089 modules in the hindgut microbiome, as well as 42 modules in the host (three in the transcriptome) (Fig. 2; Additional file 6: Table S6). Furthermore, enzymatic activities were functionally predicted by Hotpep, and the data showed that abundant peroxidases (AA2, EC 1.11.1.13, EC 1.11.1.14) were functionally identified in the microbiome, including 80 modules in the midgut and 53 modules in the hindgut. Additionally, four types of LDA enzymes were also functionally identified in the microbiome, including 44 aryl-alcohol oxidases (AA3, EC 1.1.3.7), 130 cellobiose dehydrogenases (AA3, EC 1.1.99.18), 54 vanillyl-alcohol oxidases (AA4, EC 1.1.3.38), and 51 p-benzoquinone reductases (AA6, EC 1.6.5.6) (Fig. 3; Additional file 8: Table S7). Notably, no laccase was functionally identified in either the PBL host or the gut microbiome, and neither LMEs nor LDA enzymes were identified in the PBL gut transcriptomes. Following the partial modification and degradation of lignins, celluloses and hemicelluloses present in the plant biomass are released and can be attacked by a series of enzymes.

Cellulose is a linear polymer of β-D-glucose in which glucose units are linked together by β-1,4-glycosidic bonds. The depolymerization process of celluloses is as follows: first, endoglucanases (EC 3.2.1.4) randomly attack cellulose fibrils, which reveals sites for subsequent attack by exoglucanases; then, exoglucanases (EC 3.2.1.91 and EC 3.2.1.176), also known as cellobiohydrolases, remove monomers and dimers from the reducing/nonreducing ends of the glucan chain; and finally, β-glucosidases (EC 3.2.1.21) hydrolyze glucose dimers and, in some cases, cellulose-oligosaccharides to glucose [7]. In this work, a total of 30 CAZy families known to exhibit cellulase activity were identified in the PBL holobiont, corresponding to 11,578 modules, including 11,545 modules in the microbiome and 33 modules in the host (16 in the transcriptome). Among the 30 CAZy families, sixteen GH families (GH1, GH3, GH4, GH5, GH8, GH12, GH16, GH26, GH30, GH31, GH39, GH44, GH45, GH51, GH74, and GH116) contained both hemicellulases and cellulases; fourteen CAZy families, including two AA families (AA9 and AA10) and twelve GH families (GH6, GH9, GH17, GH48, GH55, GH64, GH81, GH94, GH124, GH128, GH131, and GH144), contained only cellulases. They were all present in the microbiome, while only six GH families (GH1, GH9, GH16, GH30, GH31, and GH116) were present in the gut transcriptome (Fig. 2; Additional file 6: Table S6). Among them, nine GH families (GH5, GH6, GH8, GH9, GH12, GH26, GH44, GH45, and GH51) were functionally identified as endoglucanases (EC 3.2.1.4), corresponding to 168 modules in the midgut microbiome and 349 modules in the hindgut microbiome, and only one module (GH9) was present in the gut transcriptome. Two GH families (GH1 and GH3) were functionally identified as β-glucosidases (EC 3.2.1.21), corresponding to 313 modules in the midgut microbiome and 1,173 modules in the hindgut microbiome, and only two GH1 genes were present in the gut transcriptome. Three GH families (GH5, GH6 and GH48) were functionally identified as cellobiohydrolases (EC 3.2.1.91 and EC 3.2.1.176), corresponding to 14 modules in the midgut microbiome and 15 modules in the hindgut microbiome, respectively. Furthermore, two LPMO families (AA9 and AA10) were functionally identified as oxidoreductases corresponding to 30 modules in the microbiome, demonstrating an alternative cellulose degradation strategy present in the PBL gut microbiome. In addition, one GH family (GH94), which is known as cellobiose phosphorylase (EC 2.4.1.20), corresponding to 47 modules was functionally identified in the microbiome (Fig. 3; Additional file 8: Table S7).

Hemicellulose is a polysaccharide formed from monomeric sugars and sugar acids linked together by β-1,4- and β-1,3-glycosidic bonds [47]. Therefore, compared with that of cellulose, the degradation of hemicellulose requires a more extensive enzymatic arsenal. The current data showed that hemicellulases were the most abundant lignocellulose-degrading CAZymes in the PBL holobiont, representing 87.24%, 89.47% and 70.27% of the identified LDMs in the midgut microbiome, hindgut microbiome and host, respectively. Among these hemicellulase families, fifty-seven (ten CE families and 47 GH families) were identified in the microbiome, and twelve (one CE family and 11 GH families) were identified in the gut transcriptome (Fig. 2; Additional file 6: Table S6). Xylan is the main carbohydrate in hemicellulose. The functional prediction of the enzymatic activities demonstrated a multifunctional xylanolytic enzyme system present in the PBL microbiome. As shown in Fig. 3, xylan hydrolysis is involved in several enzymatic hydrolysis processes. First, the xylan backbone is randomly cleaved by endoxylanase (EC 3.2.1.8), and in the PBL microbiome, 1,101 endoxylanase modules were identified, with 79.65% of them present in the hindgut microbiome. Then, the xylose polymer is broken down to its monomeric form by the action of β-xylosidase (exoxylanase, EC 3.2.1.37); 272 modules from the midgut microbiome and 1,058 modules from the hindgut microbiome were predicted to have exoxylanase catalytic activity. In the PBL microbiome, various enzymes with debranching activity were also predicted, which are essential for xylan hydrolyzation [9, 48], including 137 α-glucuronidases (EC 3.2.1.139) and 460 acetylxylan esterases (EC 3.1.1.72) responsible for removal of the acetyl and phenolic side branches, 806 α-L-arabinofuranosidases (EC 3.2.1.55) catalyzing the removal of side groups, 152 feruloyl esterases (EC 3.1.1.73) cleaving the ester bonds

present on xylan, and 672 α-galactosidases (EC 3.2.1.22) catalyzing hydrolysis of the terminal α-galactosyl moieties. Additionally, in softwood, mannan is the major component of hemicellulose. In the PBL gut microbiome, a total of 181 endomannosidase (EC 3.2.1.78) and 121 exomannosidase (EC 3.2.1.25) modules were also identified for mannan hemicellulose degradation. For the PBL host, only one α-galactosidase module (GH27) and one exomannosidase module (GH2) were identified in the gut transcriptome (Additional file 8: Table S7).

Taxonomic origin of lignocellulose degradation-related genes from the PBL gut microbiome

Genes encoding lignocellulose-degrading CAZymes and lignocellulose-binding modules were searched against the NCBI NR protein database to assign taxonomic origin. In total, 12,280 genes from the midgut microbiome and 38,287 genes from the hindgut microbiome were analyzed. The results showed that 11,095 (90.35%) genes from the midgut microbiome and 33,132 (86.54%) genes from the hindgut microbiome were assigned to prokaryotic species.

When profiling the taxonomic origin, the data indicated that the PBL hindgut could selectively enrich lignocellulose-degrading microbial species. In the PBL midgut, *Firmicutes* contributed only 6.17% of lignocellulose degradation-related genes, although *Firmicutes* accounted for 52.02% of the midgut bacteria flora. However, in the hindgut, 60.46% of lignocellulose degradation-related genes were contributed by *Firmicutes*, which was higher than its relative abundance (40.55%) in the microbial community (Fig. 1A). In addition to *Firmicutes*, *Bacteroidetes* was also worth noting for its contribution of lignocellulose degradation-related genes. In the midgut, this phylum contributed 11.86% of lignocellulose degradation-related genes, although its relative abundance in the microbial community was rare (0.11%); in the hindgut, the relative abundance of *Bacteroidetes* significantly increased to 24.23%, and *Bacteroidetes* contributed 16.76% of lignocellulose degradation-related genes. At the family level, the main contributors were also concentrated in several *Firmicutes* families, including Ruminococcaceae, Lachnospiraceae, Paenibacillaceae, and Clostridiaceae, as well as one *Bacteroidetes* family, Bacteroidaceae, which contributed 57.28%, 48.16%, 52.18%, and 57.63% of cellulases, hemicellulases, CAZymes containing both cellulases and hemicellulases, and lignocellulose-binding module encoding genes, respectively (Fig. 1B). However, these families contributed very little to the lignocellulose degradation-related genes in the midgut (lower than 2.57%).

When focused on sequence novelty, our results indicated that most of the lignocellulose degradation-related genes from the PBL gut microbiome were novel. Sequence identity analysis indicated that only 5.48% of the predicted lignocellulose degradation-related proteins were highly conserved and shared more than 90% identity with the best-hit homologs in the NCBI NR database. Regarding the gut compartment, the lignocellulose-degrading enzymes and lignocellulose-binding modules enriched in the hindgut were more novel than those in the midgut. The Wilcoxon test demonstrated that the amino acid identity of LDM proteins in the hindgut was significantly lower than that in the midgut ($p < 0.001$, Fig. 1C), suggesting the potential for discovering valuable and novel enzyme resources for lignocellulose degradation from the PBL hindgut.

Metagenomic bin reconstruction and lignocellulolytic potential

To further analyze the lignocellulolytic potential of the community at the individual microbial species level, metagenomic contig binning was performed to reconstruct the genomes from PBL gut microbial communities. In this investigation, a total of 2,526 metagenomic bins were obtained (Additional file 9: Table S8), and 48.48% and 54.61% of the midgut and hindgut metagenomic reads mapped back to these bins, respectively. Completeness assessment analysis showed that 1,110 bins were substantially complete (≥ 70% completeness) and that 574 bins were near complete (≥ 90%). Then, 164 substantially complete bins with low contamination levels (≤ 5%) [33] and

high relative abundance were selected for subsequent analyses (Additional file 10: Fig. S1; Additional file 11: Table S9). The phylogenetic reconstruction of these 164 bins indicated that *Firmicutes* was the dominant phylum and comprised 56.10% (N=92) of all bins, followed by *Bacteroidetes* (N=21), *Actinobacteria* (N=13), and *Proteobacteria* (N=9). To evaluate the lignocellulose degradation potential of the 164 selected bins, the lignocellulose-degrading enzyme and lignocellulose-binding module of bins were analyzed. The data showed that most of these bins (N=156) possessed LDMs, and 71 of them were specifically notable because of their possibility of independent lignocellulose degradation, based on the possession of endo-hemicellulases, exo-hemicellulases and debranching enzymes as well as endoglucanases and β-glucosidases.

Regarding LDMs, the data showed that the distribution of exoglucanases was not universal, as these enzymes were detected in only 41 bins. In contrast, numerous endoglucanases were identified in the 71 bins, including members from the GH5, GH6, GH8, GH9, GH12, GH26, GH44 and GH51 families. These data indicated that the lack of exoglucanases may be compensated by endoglucanases [49], despite their inefficiency against crystalline cellulose. Several bins were predicted to have strong lignocellulose-degrading potential based on the composition of their LDMs. For instance, Bin-1461, Bin-2250, Bin-1063 and Bin-1473 were noteworthy for possessing LPMOs; Bin-2127 and Bin-1177 were noteworthy for possessing the largest number of cellulases and hemicellulases (Fig. 4). Furthermore, none of the 71 bins could degrade lignin due to the lack of essential LMEs for lignin degradation.

Regarding the taxonomic assignment of these bins with independent lignocellulose degradation capability, the results indicated that more than half of these bins could be novel species. Among the 71 bins, 36 bins were assigned only to the above-species level, including 19 at the phylum level, one at the order level, eight at the family level, and eight at the genus level, suggesting the presence of valuable novel microbial species resources in the PBL gut microbiota for lignocellulosic biomass conversion.

Discussion

The highly efficient lignocellulose degradation mechanism of the larvae of the saprophagous insect PB has recently gained increasing research interest due to their potential not only in farming edible insects but also in biotechnological applications. Recently, to better promote PBLs biological research and understand the genetic basis of PBLs biological characteristics, we sequenced and assembled the first PB genome [21]. The subsequent gene annotation showed that PBLs are not able to complete the process of lignocellulose degradation by themselves, indicating that the highly efficient lignocellulose degradation in PBLs may be attributed to their microbial symbionts. In this work, we investigated for the first time the complete enzyme repertoire for lignocellulose degradation on the scale of holobiont in PBLs. Combining the host gut transcriptome with the gut metagenomic data, we established a complete gut reference gene catalog that allowed us to further characterize both endogenous and microbial enzymes associated with the breakdown of lignocellulose in the PBL gut.

Overall, the PBL holobiont was assembled like a mini automatic production line for lignocellulose degradation (Fig. 5). First, PBL feeding habits drive the production line. In nature, scarab larvae such as PBLs are attracted to carbon dioxide (CO_2) [50], which drives saprophagous PBLs to feed on decaying organic matter. This feeding is beneficial to many aspects of lignocellulose degradation in the PBL holobiont: a) the PBL chews and crushes lignocellulosic biomass, which may reduce recalcitrance of the substrate and allow PBL to achieve greater lignocellulose degradation efficiency [51]; b) the PBL ingests a large number of lignocellulose-decomposing bacteria from the decaying organic matters, most of which will be killed and hydrolyzed in the midgut, providing nutrients for PBL development [52], while some surviving from the midgut will contribute to the hindgut

lignocellulosic bacteria flora [53]. Beyond feeding habits, a strong alkaline environment in the PBL midgut facilitates the solubility of organic polymers. Organic carbon analysis of intestinal contents showed that more than half of the biomass was solubilized in the highly alkaline midgut (Additional file 12: Table S10), which was capable of rendering organic components accessible to enzymatic digestion [6]. The hindgut of scarab larvae is considered to be the primary site for lignocellulose digestion, analogous to the rumen of herbivore ruminants and known as the “fermentation chamber” [54, 55]. In the present study, the annotation of metagenomic data showed a high abundance and diversity of microbial lignocellulose degradation-related enzymes in the PBL hindgut. Then, the organic matter fermented in the hindgut is dehydrated in the rectum, forming granular feces, and excreted. Therefore, the unique feeding and digestion process of PBL demonstrates that the PBL holobiont can be used as a valuable research model to study the degradation and utilization of lignocellulose.

The PBL reference gene catalog and gut content analysis data illustrate a unique teamwork between the PBL and its gut bacterial flora. In some holobionts, lignocellulose degradation is achieved via teamwork by the host and symbionts, such as in wood-feeding termites [56, 57] and omnivorous-feeding pill bugs [58], in which the lignocellulose degradation enzymatic cocktail is complemented by enzymes produced by both the host and the symbionts. However, in the PBL holobiont, the symbiont plays the major role in lignocellulose degradation processes by providing a complete and abundant enzyme repertoire. The host provides only a very limited number of enzymes involved in the depolymerization of lignocelluloses, but it poses a functional complement by providing lignin pretreatment processes. Pretreatment is the first step in the lignocellulosic material utilization process to promote access to cellulose and hemicellulose, including various physical or chemical pretreatment approaches [59]. Among these, physical pretreatment, such as grinding or milling, can reduce the particle size of materials, and chemical alkali pretreatment can lead to the separation of structural linkages between lignin and carbohydrates and disruption of the lignin structure, which have been widely applied in the pulp industry [60, 61]. In our model, the PBL host was able to provide both physical and chemical pretreatment options by evolving strong mouthparts as well as a strong alkaline environment in the midgut, which could complement the lack of laccases [62] in the PBL holobiont and promote the subsequent enzymatic hydrolysis of cellulose and hemicellulose in the hindgut. Analysis of the consumption of organic carbon during the PBL conversion process confirmed the functional complementation of the PBL host (Additional file 12: Table S10).

Regarding the enzymes, a broad array of cellulases, hemicellulases and ligninases for lignocellulosic biomass degradation were identified in the hindgut microbiome, which were at equal levels compared with that of cow rumen metagenome [54]. Although laccase, an important lignin-degrading enzyme [62], was lacking in both the PBL hindgut and rumen metagenomes, the number of peroxidases and LDA enzymes in the PBL hindgut were higher than that in the rumens of cow [54] and camel [55] (Additional file 13: Table S11), demonstrating the capability of lignin modification in the PBL hindgut microbiome. Microbial source analysis demonstrated that these LDM-encoding genes in the PBL hindgut were similar to those in rumens and were mainly contributed by two phyla, *Firmicutes* and *Bacteroidetes*, although with different levels of contribution [55, 63]. Members of these two phyla are known as potent lignocellulose degraders, and their association with lignocellulose degradation has been well established [64]. Furthermore, our data indicated that most of the LDM sequences enriched by the PBL symbiont were novel, sharing lower than 90% identity with the best-hit homologs in the NCBI NR database, similar to cow rumen [54]. These data point to valuable and novel genetic resources for carbohydrate degradation in the PBL microbiome.

Regarding the microbiota, further taxonomic assignment at the individual microbial species level proved the existence of diverse and novel lignocellulosic microflora present in the PBL gut. A total of 156 high-quality

metagenomic bins showing lignocellulolytic potential were reconstructed in this work, and as expected, 70% of these bins were associated with species belonging to *Firmicutes* and *Bacteroidetes*. Among these, 71 bins were identified with independent lignocellulose degradation capability and most (50.70%) of these genomes represent previously unsequenced strains and species, demonstrating discovery of novel bacterial species associated with lignocellulose degradation in the PBL microbiota. To assess the reliability of the metagenome binning results in this work, for 35 bins resolvable to the species level, we collected the public genomes of the same species from the NCBI database and analyzed the potential of lignocellulose degradation for these published genomes. The data indicated that the LDM genes were also present in all these published genomes (Additional file 14: Table S12). Some of the species these bins assigned to were isolated from known lignocellulolytic organisms, such as ruminants or termites, and have been proven to have lignocellulose degradation capability. For instance, Bin-2127 possessed the highest number of LDMs (273 LDMs) and was identified as a strain of *Bacteroides faecis*, a species isolated from human feces that has been proven to be a decomposer of various mono/polysaccharides [65], and the annotation of the published *B. faecis* strain genome (GCA_000226135.1) revealed 189 LDM genes. Additionally, another *Bacteroidetes* bin (Bin-821) was assigned to *Sporocytophaga myxococcoides*, a species that has been regarded as a highly efficient carbohydrate metabolizer possessing a wide array of cellulolytic enzymes [66]. The annotation of its public genomes (GCA_000426725.1) revealed complete endoglucanases, exoglucanases and β-glucosidases for cellulose degradation. For species from *Firmicutes*, Bin-1076 was assigned to *Clostridium cellulosi*, which is a thermophilic cellulolytic microbe that produces a series of extracellular GHs and is capable of fermenting some difficult substrates, including glycogen, inulin, mannitol and sucrose, to produce acetylmethylcarbinol [67]; Bin-1067 was assigned to *Clostridium sartagoforme*, a species that was isolated from cow dung compost and could directly utilize various carbon sources to produce hydrogen [68]; Bin-526 was assigned to *Herbinix luporum*, a species from a thermophilic biogas plant, which is able to degrade crystalline cellulose [69]; and Bin-2333 was assigned to *Ruminococcus bromii*, which plays a significant role in the breakdown of plant polysaccharides, such as resistant starch, in the human gut [70]. These bins and corresponding public genomes also possessed a high number of cellulase and hemicellulase genes (Additional file 14: Table S12). Overall, the consistency of LDM genes between the reconstructed bins and the corresponding species' genomes further illustrated the representativeness of the PBL metagenomic features we demonstrated in this study, as well as the research value and application prospects of the PBLs lignocellulose degradation model.

Conclusions

In summary, a comprehensive reference catalog of gut microbial genes and host gut transcriptomic genes was first established for PBLs in this work. We characterized a gene repertoire comprising highly abundant and diversified lignocellulose-degrading enzymes and demonstrated that there is unique teamwork between PBLs and their gut bacterial flora for efficient lignocellulose degradation. The PBLs selectively enrich lignocellulose-degrading microbial species mainly from *Firmicutes* and *Bacteroidetes*, which are capable of producing a broad array of cellulases and hemicellulases, thus playing a major role in lignocellulosic biomass degradation. The PBL hosts provide lignin pretreatment processes via their evolved strong mouthparts as well as a strong alkaline environment in the midgut, which thus pose functional complementation. In addition, most of the LDMs sequences in the PBL microbiome were novel. And 71 high-quality genome bins with independent cellulose and hemicellulose degradation capability were recovered from the PBL gut microbiome, most of which were novel species. This study provides valuable genetic resources for further discovery of novel enzymes and microbial candidates not only to enable improvement of PBL feeding models with higher feed efficiency but also for industrial applications such as biofuel and biocatalysts.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

The datasets generated and/or analyzed during the current study are available in the NCBI SRA repository under accession numbers provided in Additional file 1.

Competing interests

The authors declare that they have no competing interests.

Funding

Funds to support this work were provided by the National Natural Science Foundation of China (No. 32070511 and No. 31972336).

Authors' contributions

CS conceived and designed the experiments. KW conducted the genomics experiments, analyzed and compiled data. KW and CS wrote the manuscript. PG analyzed organic carbon content in the PBL midgut and hindgut. CL and LG provided the insect and collected samples. JZ and CS oversaw these experiments and data analysis, and edited the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

[1] Diamond J. Evolution, consequences and future of plant and animal domestication.Nature. 2002; 418**:** 700-7.

[2] Ham YK, Kim SW, Song DH, Kim HW, Kim IS. Nutritional composition of white-spotted flower chafer (*Protaetia brevitarsis*) larvae produced from commercial insect farms in korea.Food Sci Anim Resour. 2021; 41**:** 416-27.

[3] Li Y, Fu T, Geng L, Shi Y, Chu H, Liu F, et al. *Protaetia brevitarsis* larvae can efficiently convert herbaceous and ligneous plant residues to humic acids.Waste Manage. 2019; 83**:** 79-82.

[4] Wei P, Li Y, Lai D, Geng L, Liu C, Zhang J, et al. *Protaetia brevitarsis* larvae can feed on and convert spent mushroom substrate from Auricularia auricula and Lentinula edodes cultivation.Waste Manage. 2020; 114**:** 234-9.

[5] Huang S, Zhang H, Marshall S, Jackson TA. The scarabgut: A potential bioreactor for bio-fuel production.Insect Sci. 2010; 17**:** 175-83.

[6] Lemke T, Stingl U, Egert M, Friedrich MW, Brune A. Physicochemical conditions and microbial activities in the highly alkaline gut of the humus-feeding larva of *Pachnoda ephippiata* (Coleoptera: Scarabaeidae).Appl Environ Microb. 2003; 69**:** 6650-8.

[7] Andlar M, Rezić T, Marđetko N, Kracher D, Ludwig R, Šantek B. Lignocellulose degradation: an overview of fungi and fungal enzymes involved in lignocellulose degradation.Eng Life Sci. 2018; 18**:** 768-78.

[8] Silva JP, Ticona ARP, Hamann PRV, Quirino BF, Noronha EF. Deconstruction of lignin: from enzymes to microorganisms.Molecules. 2021; 26**:** 2299.

[9] Lange L. Fungal enzymes and yeasts for conversion of plant biomass to bioenergy and high-value products.Microbiol Spectr. 2017; 5.

[10] Sabbadin F, Hemsworth GR, Ciano L, Henrissat B, Dupree P, Tryfona T, et al. An ancient family of lytic polysaccharide monooxygenases with roles in arthropod development and biomass digestion.Nat Commun. 2018; 9**:** 756.

[11] Couturier M, Ladevèze S, Sulzenbacher G, Ciano L, Fanuel M, Moreau C, et al. Lytic xylan oxidases from wood-decay fungi unlock biomass degradation.Nat Chem Biol. 2018; 14**:** 306-10.

[12] Hemsworth GR, Johnston EM, Davies GJ, Walton PH. Lytic polysaccharide monooxygenases in biomass conversion.Trends Biotechnol. 2015; 33**:** 747-61.

[13] Houfani AA, Anders N, Spiess AC, Baldrian P, Benallaoua S. Insights from enzymatic degradation of cellulose and hemicellulose to fermentable sugars– a review.Biomass Bioenerg. 2020; 134**:** 105481.

[14] Bredon M, Herran B, Bertaux J, Grève P, Moumen B, Bouchon D. Isopod holobionts as promising models for lignocellulose degradation.Biotechnol Biofuels. 2020; 13**:** 49.

[15] Brune A. Symbiotic digestion of lignocellulose in termite guts.Nat Rev Microbiol. 2014; 12**:** 168-80.

[16] Zeng W, Liu B, Zhong J, Li Q, Li Z. A natural high-sugar diet has different effects on the prokaryotic community structures of lower and higher termites (Blattaria).Environ Entomol. 2020; 49**:** 21-32.

[17] Su L, Yang L, Huang S, Su X, Li Y, Wang F, et al. Comparative gut microbiomes of four species representing the higher and the lower termites.J Insect Sci. 2016; 16**:** 97.

[18] Wang N, Wang W, Jiang Y, Dai W, Li P, Yao D, et al. Variations in bacterial taxonomic profiles and potential functions in response to the gut transit of earthworms (*Eisenia fetida*) feeding on cow manure.Sci Total Environ. 2021; 787**:** 147392.

[19] Kataev VY, Sleptsov, II, Martynov AA, Aduchiev BK, Khlopko YA, Miroshnikov SA, et al. Data on rumen and faeces microbiota profiles of Yakutian and Kalmyk cattle revealed by high-throughput sequencing of 16S rRNA gene amplicons.Data Brief. 2020; 33**:** 106407.

[20] Tian XY, Song FP, Zhang J, Liu RM, Zhang XP, Duan JY, et al. Diversity of gut bacteria in larval *Protaetia brevitarsis*(Coleoptera: Scarabaedia) fed on corn stalk.Acta Entomologica Sinica 2017; 60**:** 632-41.

[21] Wang K, Li P, Gao Y, Liu C, Wang Q, Yin J, et al. *De novo* genome assembly of the white-spotted flower chafer (*Protaetia brevitarsis*).GigaScience. 2019; 8**:** giz019.

[22] Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor.Bioinformatics. 2018; 34**:** 884-90.

[23] Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner.Bioinformatics. 2013; 29**:** 15-21.

[24] Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads.Nat Biotechnol. 2015; 33**:** 290-5.

[25] Bolger AM, Marc L, Bjoern U. Trimmomatic: a flexible trimmer for Illumina sequence data.Bioinformatics. 2014; 30**:** 2114-20.

[26] Beghini F, McIver LJ, Blanco-Míguez A, Dubois L, Asnicar F, Maharjan S, et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3.elife. 2021; 10**:** e65088.

[27] Li D, Liu CM, Luo R, Kunihiko S, Tak-Wah L. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct *de Bruijn*graph.Bioinformatics. 2015; 31**:** 1674-6.

[28] Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.ArXiv. 2013; 1303**:** 3997.

[29] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools.Bioinformatics. 2009; 25**:** 2078-9.

[30] Wu YW, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets.Bioinformatics. 2016; 32**:** 605-7.

[31] Kang DD, Li F, Kirton E, Thomas A, Wang Z. MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies.PeerJ. 2019; 7**:** e7359.

[32] Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication.ISME J. 2017; 11**:** 2864-8.

[33] Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes.Genome Res. 2015; 25**:** 1043-55.

[34] Stewart RD, Auffret MD, Warr A, Walker AW, Roehe R, Watson M. Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery.Nat Biotechnol. 2019; 37**:** 953-61.

[35] Stewart RD, Auffret MD, Snelling TJ, Roehe R, Watson M. MAGpy: a reproducible pipeline for the downstream analysis of metagenome-assembled genomes (MAGs).Bioinformatics. 2019; 35**:** 2150-2.

[36] Asnicar F, Thomas AM, Beghini F, Mengoni C, Manara S, Manghi P, et al. Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0.Nat Commun. 2020; 11**:** 2500.

[37] Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation.Nucleic Acids Res. 2021; 49**:** 293-6.

[38] Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification.BMC Bioinformatics. 2010; 11**:** 119.

[39] Seemann T. Prokka: rapid prokaryotic genome annotation.Bioinformatics. 2014; 30**:** 2068-9.

[40] Uritskiy GV, DiRuggiero J, Taylor J. MetaWRAP-a flexible pipeline for genome-resolved metagenomic data analysis.Microbiome. 2018; 6**:** 158.

[41] Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013.Nucleic Acids Res. 2014; 42**:** 490-5.

[42] Zhang H, Tanner Y, Huang L, Sarah E, Wu P, Yang Z, et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation.Nucleic Acids Res. 2018; 46**:** 95-101.

[43] Busk PK, Pilgaard B, Lezyk MJ, Meyer AS, Lange L. Homology to peptide pattern for annotation of carbohydrate-active enzymes and prediction of function.BMC Bioinformatics. 2017; 18**:** 214.

[44] Buchfink B, Reuter K, Drost HG. Sensitive protein alignments at tree-of-life scale using DIAMOND.Nat Methods. 2021; 18**:** 366-8.

[45] Huson DH, Beier S, Flade I, Górska A, El-Hadidi M, Mitra S, et al. MEGAN community edition - interactive exploration and analysis of large-scale microbiome sequencing data.PLoS Comput Biol. 2016; 12**:** e1004957.

[46] Geng A, Cheng Y, Wang Y, Zhu D, Le Y, Wu J, et al. Transcriptome analysis of the digestive system of a wood-feeding termite (*Coptotermes formosanus*) revealed a unique mechanism for effective biomass degradation.Biotechnol Biofuels. 2018; 11**:** 24.

[47] Shallom D, Shoham Y. Microbial hemicellulases.Curr Opin Microbiol. 2003; 6**:** 219-28.

[48] Dutta S, Wu KCW. Enzymatic breakdown of biomass: enzyme active sites, immobilization, and biofuel production.Green Chem. 2014; 16**:** 4615-26.

[49] Calderón-Cortés N, Quesada M, Watanabe H, Cano-Camacho H, Oyama K. Endogenous plant cell wall digestion: a key mechanism in insect evolution.Annu Rev Ecol Evol S. 2012; 43**:** 45-71.

[50] Kojima W. Attraction to carbon dioxide from feeding resources and conspecific neighbours in larvae of the rhinoceros beetle *Trypoxylus dichotomus*.PLoS One. 2015; 10**:** e0141733.

[51] Rajeswari G, Jacob S, Chandel AK, Kumar V. Unlocking the potential of insect and ruminant host symbionts for recycling of lignocellulosic carbon with a biorefinery approach: a review.Microb Cell Fact. 2021; 20**:** 107.

[52] Vallet-Gely I, Lemaitre B, Boccard F. Bacterial strategies to overcome insect defences.Nat Rev Microbiol. 2008; 6**:** 302-13.

[53] Su L, Yang L, Huang S, Li Y, Su X, Wang F, et al. Variation in the gut microbiota of termites (*Tsaitermes ampliceps*) against different diets.Appl Biochem Biotech. 2017; 181**:** 32-47.

[54] Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G, et al. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen.Science. 2011; 331**:** 463-7.

[55] Gharechahi J, Salekdeh GH. A metagenomic analysis of the camel rumen's microbiome identifies the major microbes responsible for lignocellulose degradation and fermentation.Biotechnol Biofuels. 2018; 11**:** 216.

[56] Poulsen M, Hu H, Li C, Chen Z, Xu L, Otani S, et al. Complementary symbiont contributions to plant decomposition in a fungus-farming termite.P Natl Acad Sci USA. 2014; 111**:** 14500-5.

[57] Ni J, Tokuda G. Lignocellulose-degrading enzymes from termites and their symbiotic microbiota.Biotechnol Adv. 2013; 31**:** 838-50.

[58] Bredon M, Dittmer J, Noël C, Moumen B, Bouchon D. Lignocellulose degradation at the holobiont level: teamwork in a keystone soil invertebrate.Microbiome. 2018; 6**:** 162.

[59] Galbe M, Wallberg O. Pretreatment for biorefineries: a review of common methods for efficient utilisation of lignocellulosic materials.Biotechnol Biofuels. 2019; 12**:** 294.

[60] Chandra RP, Chu Q, Hu J, Zhong N, Lin M, Lee JS, et al. The influence of lignin on steam pretreatment and mechanical pulping of poplar to achieve high sugar recovery and ease of enzymatic hydrolysis.Bioresource Technol. 2016; 199**:** 135-41.

[61] Rojas OJ, Hubbe MA. The dispersion science of papermaking.J Disper Sci Technol. 2005; 25**:** 713-32.

[62] Gupta R, Mehta G, Khasa YP, Kuhad RC. Fungal delignification of lignocellulosic biomass improves the saccharification of cellulosics.Biodegradation. 2011; 22**:** 797-804.

[63] Svartström O, Alneberg J, Terrapon N, Lombard V, De Bruijn I, Malmsten J, et al. Ninety-nine *de novo* assembled genomes from the moose (*Alces alces*) rumen microbiome provide new insights into microbial plant biomass degradation.ISME J. 2017; 11**:** 2538-51.

[64] Jami E, Israel A, Kotser A, Mizrahi I. Exploring the bovine rumen bacterial community from birth to adulthood.ISME J. 2013; 7**:** 1069-79.

[65] Kim MS, Roh SW, Bae JW. *Bacteroides faecis* sp. nov., isolated from human faeces.Int J Syst Evol Micr. 2010; 60**:** 2572-6.

[66] Taillefer M, Arntzen M, Henrissat B, Pope PB, Larsbrink J. Proteomic dissection of the cellulolytic machineries used by soil-dwelling bacteroidetes.mSystems. 2018; 3**:** e00240-18.

[67] Zhang K, Li W, Wang Y, Zheng Y, Tan F, Ma X, et al. Processive degradation of crystalline cellulose by a multimodular endoglucanase via a wirewalking mode.Biomacromolecules. 2018; 19**:** 1686-96.

[68] Zhang J, Li Y, Zheng H, Fan Y, Hou H. Direct degradation of cellulosic biomass to bio-hydrogen from a newly isolated strain *Clostridium sartagoforme*FZ11.Bioresource Technol. 2015; 192**:** 60-7.

[69] Koeck DE, Hahnke S, Zverlov VV. *Herbinix luporum* sp. nov., a thermophilic cellulose-degrading bacterium isolated from a thermophilic biogas reactor.Int J Syst Evol Micr. 2016; 66**:** 4132-7.

[70] Ze X, Duncan SH, Louis P, Flint HJ. *Ruminococcus bromii* is a keystone species for the degradation of resistant starch in the human colon.ISME J. 2012; 6**:** 1535-43.

Figures

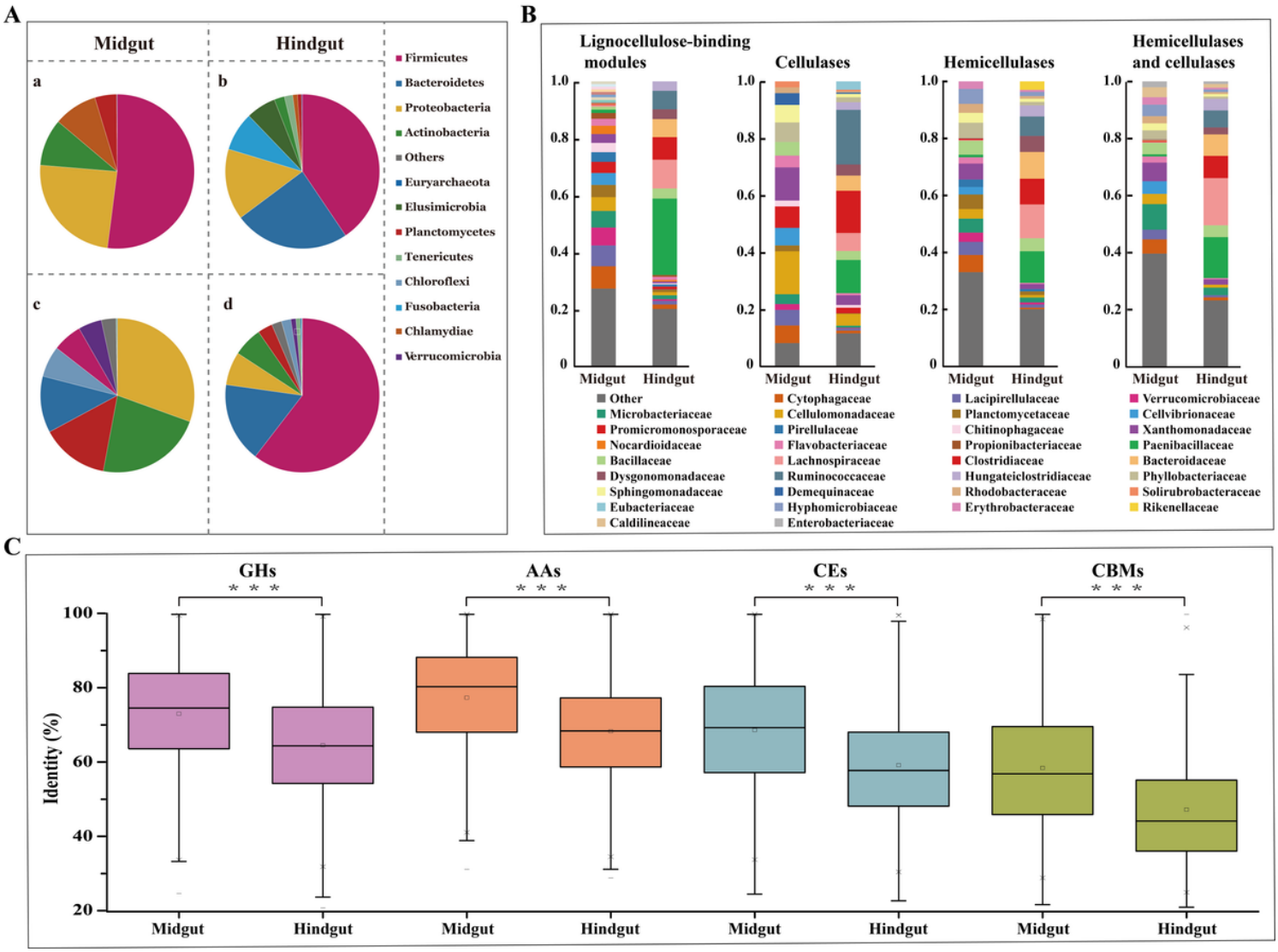

Figure 1

Taxonomic composition at the phylum level by the relative abundances of metagenomic reads in the midgut (A-a) and hindgut (A-b), or by the lignocellulose degradation-related modules predicted in the midgut (A-c) and hindgut (A-d) microbiome. Family level taxonomic distribution of the lignocellulose degradation-related modules in the midgut and hindgut microbiome (B). Similarity distribution between lignocellulose degradation-related proteins (GHs, AAs, CEs, and CBMs) and the best hit in the NCBI NR protein database (C). The percentage sequence identity of lignocellulose degradation-related proteins encoded in the midgut and hindgut microbiome is displayed by the box-plot, respectively.

LDMs	CAZy Family	Microbiome		Host	Total
		Midgut	Hindgut		
Ligninases	AA1	3	2	0	5
	AA2	157	97	0	254
	AA3	343	231	2	576
	AA4	138	156	1	295
	AA5	2	3	0	5
	AA6	229	1463	0	1692
	AA7	166	137	0	303
Cellulases	AA9	7	1	0	8
	AA10	21	23	0	44
	GH6	31	24	0	55
	GH9	58	165	1	224
	GH17	7	6	0	13
	GH48	7	5	0	12
	GH55	3	34	0	37
	GH64	9	4	0	13
	GH81	9	7	0	16
	GH94	49	462	0	511
	GH124	0	5	0	5
	GH128	20	27	0	47
	GH131	1	0	0	1
	GH144	58	268	0	326
Hemicellulases and cellulases	GH1	270	533	3	806
	GH3	558	2089	0	2647
	GH4	183	1122	0	1305
	GH5	439	1061	0	1500
	GH8	48	117	0	165
	GH12	16	4	0	20
	GH16	210	438	5	653
	GH26	62	210	0	272
	GH30	49	259	1	309
	GH31	149	604	5	758
	GH39	189	348	0	537
	GH44	8	11	0	19
	GH45	6	4	0	10
	GH51	115	479	0	594
	GH74	150	245	0	395
	GH116	41	217	1	259

LDMs	CAZy Family	Microbiome		Host	Total
		Midgut	Hindgut		
Hemicellulases	CE1	2071	2884	1	4956
	CE2	44	80	0	124
	CE3	490	884	0	1374
	CE4	930	3739	0	4669
	CE5	33	32	0	65
	CE6	146	330	0	476
	CE7	314	509	0	823
	CE12	187	385	0	572
	CE15	190	293	0	483
	CE16	24	13	0	37
	GH2	162	1402	2	1566
	GH10	205	662	0	867
	GH11	33	156	0	189
	GH27	19	109	1	129
	GH29	104	820	2	926
	GH35	54	193	1	248
	GH36	28	248	0	276
	GH38	183	806	4	993
	GH42	83	361	0	444
	GH43	612	2273	0	2885
	GH47	6	10	3	19
	GH52	1	10	0	11
	GH53	32	101	0	133
	GH54	2	1	0	3
	GH57	85	201	0	286
	GH59	6	18	0	24
	GH62	14	9	0	23
	GH67	35	130	0	165
	GH92	55	340	0	395
	GH93	34	69	0	103
	GH95	41	345	0	386
	GH97	40	173	0	213
	GH98	0	6	0	6
	GH99	33	106	0	139
	GH110	1	63	0	64
	GH113	31	41	0	72
	GH115	25	153	0	178
	GH120	33	125	0	158
	GH127	108	518	0	626
	GH134	0	1	0	1
	GH141	22	182	0	204

LDMs	CAZy Family	Microbiome		Host	Total
		Midgut	Hindgut		
Lignocellulose-binding modules	CBM1	3	1	0	4
	CBM2	82	109	0	191
	CBM3	3	47	0	50
	CBM4	87	418	0	505
	CBM6	140	612	0	752
	CBM8	5	13	0	18
	CBM9	199	674	0	873
	CBM10	2	0	0	2
	CBM11	11	20	0	31
	CBM13	73	468	1	542
	CBM16	104	546	0	650
	CBM17	0	2	0	2
	CBM22	130	573	0	703
	CBM23	31	101	0	132
	CBM27	0	28	0	28
	CBM28	1	7	0	8
	CBM29	0	6	0	6
	CBM30	28	47	0	75
	CBM31	0	3	0	3
	CBM32	430	2557	0	2987
	CBM35	165	794	0	959
	CBM36	1	9	0	10
	CBM37	38	407	0	445
	CBM39	1	4	6	11
	CBM42	8	15	0	23
	CBM44	424	345	0	769
	CBM46	3	238	0	241
	CBM47	51	45	0	96
	CBM49	0	1	0	1
	CBM51	121	290	0	411
	CBM52	0	1	0	1
	CBM54	4	839	0	843
	CBM56	26	293	0	319
	CBM59	3	4	0	7
	CBM60	15	8	0	23
	CBM61	35	223	0	258
	CBM62	7	191	0	198
	CBM63	4	6	0	10
	CBM65	0	38	0	38
	CBM67	281	974	0	1255
	CBM72	0	3	0	3
	CBM75	0	5	0	5
	CBM76	0	4	0	4
	CBM78	0	1	0	1
	CBM80	1	3	0	4
	CBM81	1	1	0	2

Figure 2

Distribution of lignocellulose-degrading CAZymes and lignocellulose-binding modules in the PBL gut gene catalog. Presented are the total numbers of CAZy modules for each family in the host gut transcriptome and in the gut metagenome.

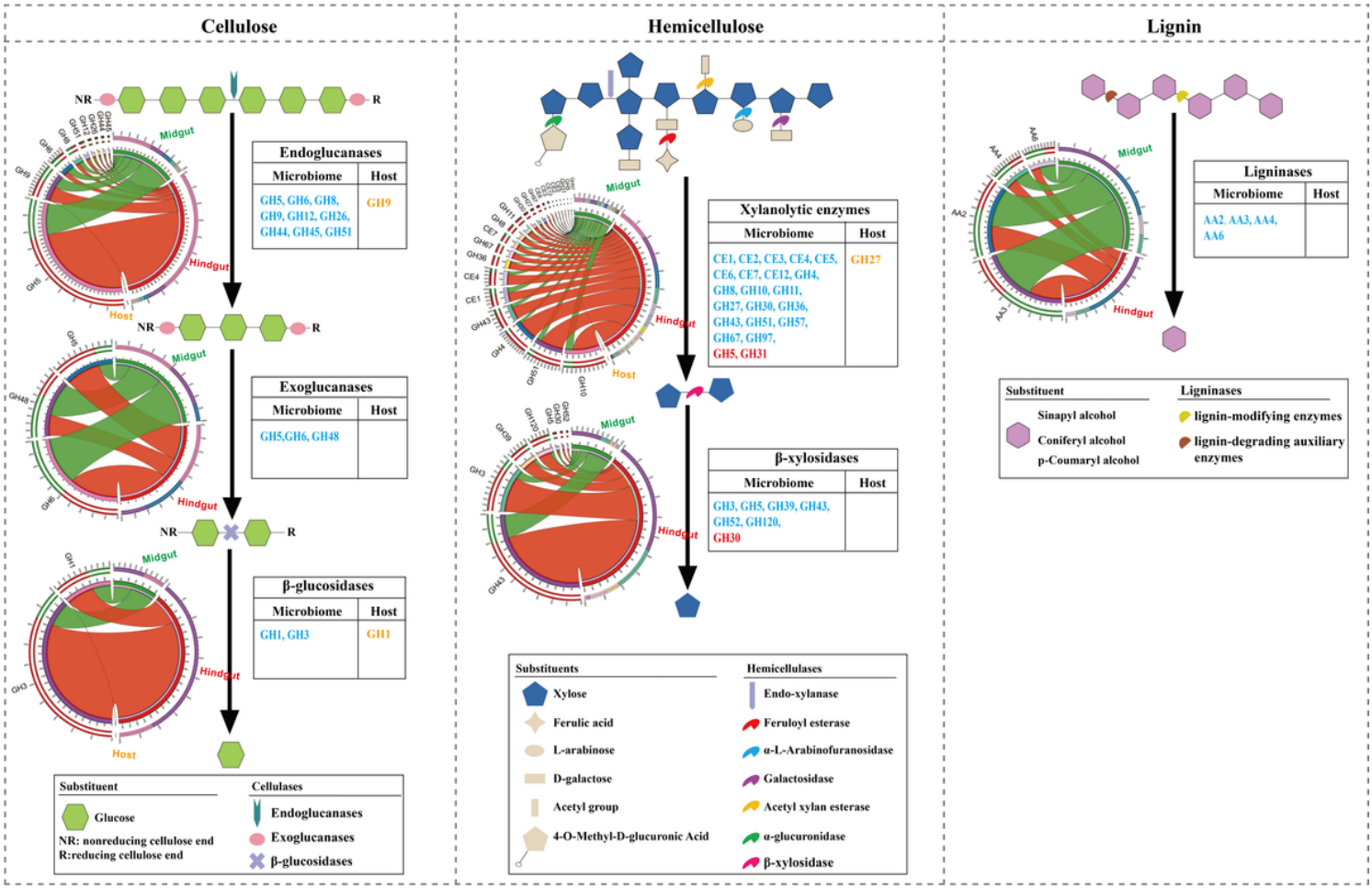

Figure 3

Cooperative model of cellulases, hemicellulases and ligninases in lignocellulose degradation in the PBL holobiont. Diagrams represent the CAZy families contributed by the host (transcriptome, orange), the midgut microbiome (green) and the hindgut microbiome (red). The CAZy families present in both the midgut microbiome and hindgut microbiome were shown in blue.

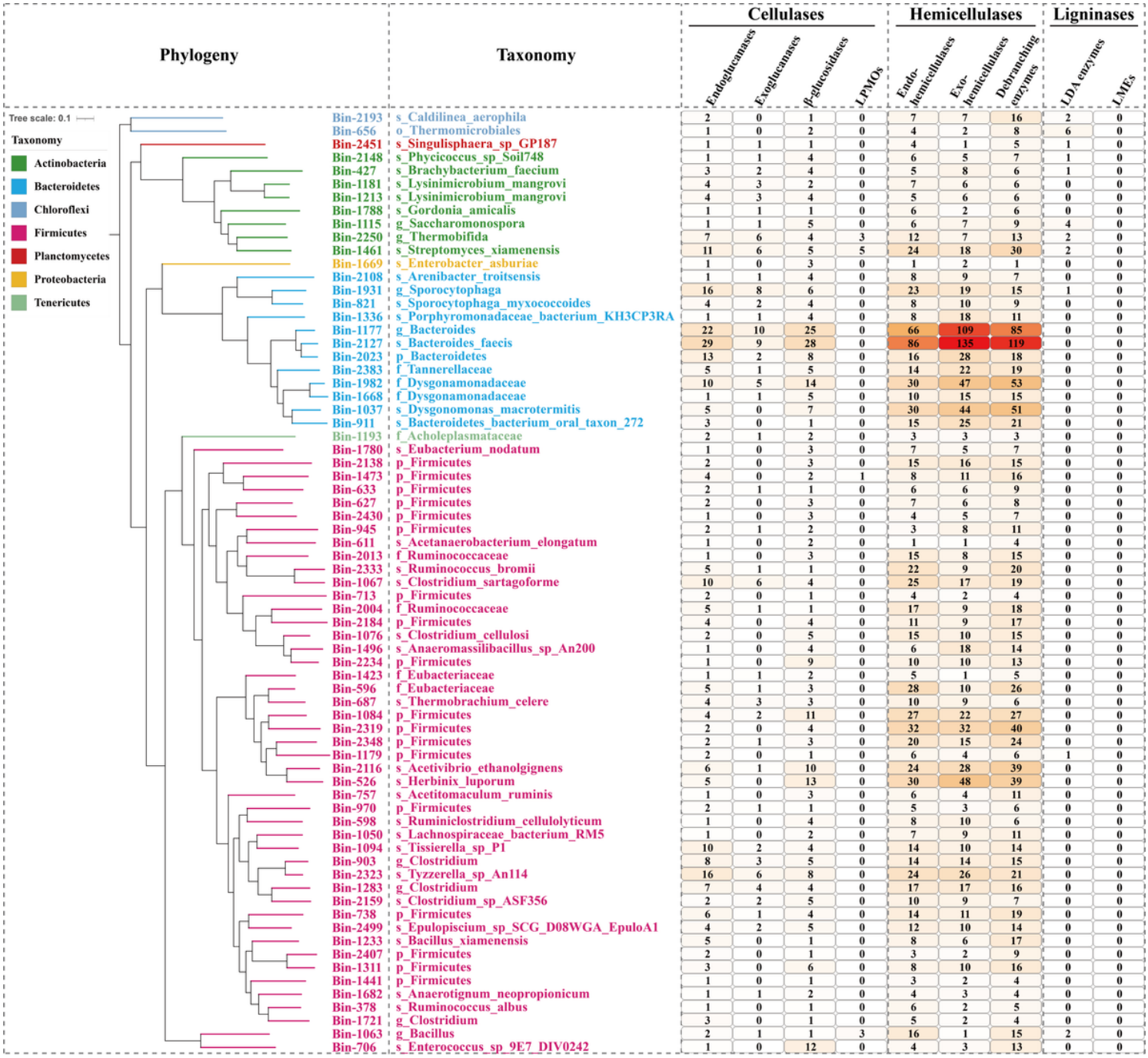

Figure 4

Phylogenetic affiliation, taxonomic assignment and metabolic potential of 71 genomic bins with independent cellulose and hemicellulose degradation capability. Branches and labels with different colors represent different phyla. Taxonomic assignment level, k_, kingdom; p_, phylum; c_, class; o_, order; f_, family; g_, genus; s_, species. The heatmap in the right depicts the number of lignocellulose-degrading CAZymes in each bin.

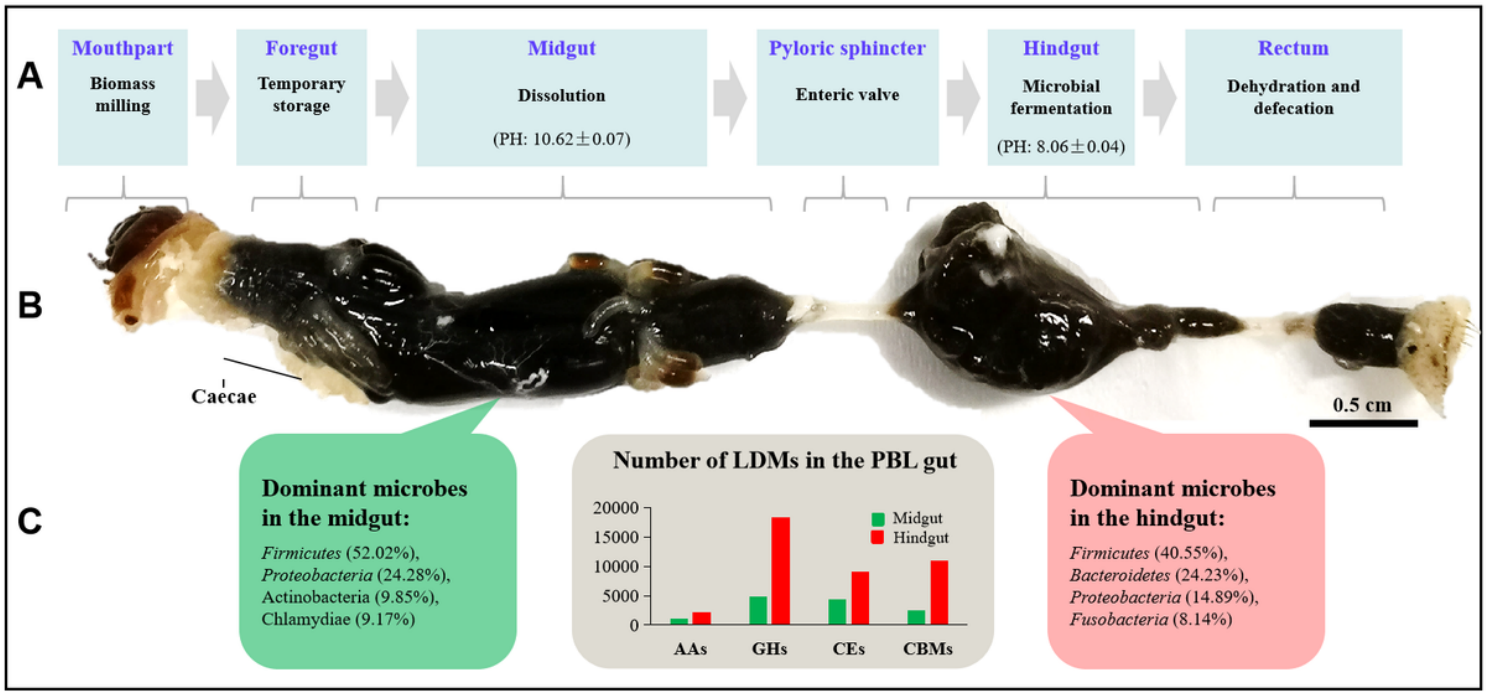

Figure 5

Structural and functional assembly of PBL digestive tract. Physicochemical properties and major functions in the lignocellulose conversion of different digestive tract compartments (A). View of the digestive tract of PBL showing relative locations of different compartments (B). Dominant microbes and the total number of LDMs in the PBL midgut and hindgut microbiome (C).

Supplementary Files

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

- Additionalfile1TableS1.xls
- Additionalfile2TableS2.xls
- Additionalfile3TableS3.xls
- Additionalfile4TableS4.xls
- Additionalfile5TableS5.xls
- Additionalfile6TableS6.xls
- Additionalfile7CAZymes.fasta
- Additionalfile8TableS7.xls
- Additionalfile9TableS8.xls
- Additionalfile10Fig.S1.png
- Additionalfile11TableS9.xls
- Additionalfile12TableS10.xls
- Additionalfile13TableS11.xls
- Additionalfile14TableS12.xls