



Carbohydrate Hydrolytic Potential and Redundancy of an Anaerobic Digestion Microbiome Exposed to Acidosis, as Uncovered by Metagenomics

Marie Bertucci,^{a,b} Magdalena Calusinska,^a Xavier Goux,^a Corinne Rouland-Lefèvre,^c Boris Untereiner,^a Pau Ferrer,^{a*} Patrick A. Gerin,^b Philippe Delfosse^{a*}

^aEnvironmental Research and Innovation Department, Luxembourg Institute of Science and Technology, Belvaux, Luxembourg

^bLaboratory of Bioengineering, Earth and Life Institute, Applied Microbiology, Université Catholique de Louvain, Louvain-la-Neuve, Belgium

^cInstitute of Ecology and Environmental Sciences, Research Institute Development, Sorbonne Universités, Bondy, France

ABSTRACT Increased hydrolysis of easily digestible biomass may lead to acidosis of anaerobic reactors and decreased methane production. Previously, it was shown that the structure of microbial communities changed during acidosis; however, once the conditions are back to optimal, biogas (initially CO₂) production quickly restarts. This suggests the retention of the community functional redundancy during the process failure. In this study, with the use of metagenomics and downstream bioinformatics analyses, we characterize the carbohydrate hydrolytic potential of the microbial community, with a special focus on acidosis. To that purpose, carbohydrate-active enzymes were identified, and to further link the community hydrolytic potential with key microbes, bacterial genomes were reconstructed. In addition, we characterized biochemically the specificity and activity of selected enzymes, thus verifying the accuracy of the *in silico* predictions. The results confirm the retention of the community hydrolytic potential during acidosis and indicate *Bacteroidetes* to be largely involved in biomass degradation. *Bacteroidetes* showed higher diversity and genomic content of carbohydrate hydrolytic enzymes that might favor the dominance of this phylum over other bacteria in some anaerobic reactors. The combination of bioinformatic analyses and activity tests enabled us to propose a model of acetylated glucomannan degradation by *Bacteroidetes*.

IMPORTANCE The enzymatic hydrolysis of lignocellulosic biomass is mainly driven by the action of carbohydrate-active enzymes. By characterizing the gene profiles at the different stages of the anaerobic digestion experiment, we showed that the microbiome retains its hydrolytic functional redundancy even during severe acidosis, despite significant changes in taxonomic composition. By analyzing reconstructed bacterial genomes, we demonstrate that *Bacteroidetes* hydrolytic gene diversity likely favors the abundance of this phylum in some anaerobic digestion systems. Further, we observe genetic redundancy within the *Bacteroidetes* group, which accounts for the preserved hydrolytic potential during acidosis. This work also uncovers new polysaccharide utilization loci involved in the deconstruction of various biomasses and proposes the model of acetylated glucomannan degradation by *Bacteroidetes*. Acetylated glucomannan-enriched biomass is a common substrate for many industries, including pulp and paper production. Using naturally evolved cocktails of enzymes for biomass pretreatment could be an interesting alternative to the commonly used chemical pretreatments.

KEYWORDS *Bacteroidetes*, biotechnology, enzymes, molecular biology, polysaccharides, recombinant-protein production

Citation Bertucci M, Calusinska M, Goux X, Rouland-Lefèvre C, Untereiner B, Ferrer P, Gerin PA, Delfosse P. 2019. Carbohydrate hydrolytic potential and redundancy of an anaerobic digestion microbiome exposed to acidosis, as uncovered by metagenomics. *Appl Environ Microbiol* 85:e00895-19. <https://doi.org/10.1128/AEM.00895-19>.

Editor Maia Kivisaar, University of Tartu

Copyright © 2019 Bertucci et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Magdalena Calusinska, magdalena.calusinska@list.lu.

* Present address: Pau Ferrer, Universitat Autònoma de Barcelona, Barcelona, Spain; Philippe Delfosse, Université du Luxembourg, Belval, Luxembourg.

Received 18 April 2019

Accepted 26 May 2019

Accepted manuscript posted online 31 May 2019

Published 18 July 2019

Anaerobic digestion (AD) of biomass (including biowaste) is a process aiming at producing biogas, i.e., a mixture of methane (CH_4), carbon dioxide (CO_2), and trace gases (i.e., H_2 , NH_3 , and H_2S) (1). Biogas is a result of the joint action of diverse microorganisms that act synergistically to decompose organic matter. In the first stage, namely hydrolysis, the organic matter (i.e., fats, proteins, and polysaccharides) is deconstructed by fermentative bacteria into soluble molecules. During the acidogenesis and acetogenesis stages, bacterial consortia convert the resulting monomers and oligomers mainly into volatile fatty acids (VFAs), including acetate, CO_2 , and H_2 . Finally, methane is produced by methanogenic archaea during the methanogenesis stage (2). Even though different environmental factors and operational conditions can lead to the dysfunctioning of anaerobic digesters, acidification caused by an accumulation of VFAs (here, referred to as acidosis) is one of the most often recorded phenomena (3, 4). It results from the decoupling of the hydrolytic and acidogenic stages (which perform too fast due to, e.g., higher multiplication rates of involved organisms under specific conditions) from the downstream acetogenesis and methanogenesis stages (which are too slow due to, e.g., slower multiplication rates of the acetogenic bacteria and methanogenic archaea) (5). As reported from previous studies, acidosis leads to a change in the microbial community structure, i.e., the taxonomic composition (4, 6). Both bacterial and archaeal communities are affected, and, as a result, methane production is slowed down and sometimes interrupted. Interestingly, following a recovery stage, the newly established community relatively quickly restarts the biogas production (3, 4). This suggests that the functional community redundancy (similar metabolic functions performed by distinct coexisting microorganisms) (7) occurs in AD during the acidosis.

Hydrolysis is often considered a bottleneck of the AD process because it either underperforms (in the case of the recalcitrant biomass, e.g., highly lignified or high crystalline cellulose content) or outperforms (fast hydrolysis of an easily digestible biomass can lead to acidosis). The enzymatic hydrolysis of lignocellulose biomass is mainly driven by the action of carbohydrate active enzymes (CAZymes) and especially by glycoside hydrolases (GHs), carbohydrate esterases (CEs), polysaccharide lyases (PLs), and other auxiliary enzymes (AAs) (8). Carbohydrate-binding modules (CBMs) are noncatalytic contiguous amino acid sequences, which bind hydrolytic enzymes to their carbohydrate substrates. They usually exist as modules within larger enzymes, but some can be independent proteins (9). Glycosyltransferases (GTs) are also classified within the CAZy database as they are involved in the biosynthesis of glycosidic bonds from phosphate-activated sugars donors. GHs hydrolyze and/or transglycosylate the glycosidic bonds, CEs hydrolyze ester bonds, and AAs are redox enzymes acting in conjunction with other CAZymes, many of which are involved in lignin degradation. PLs cleave bonds from uronic acid-containing polysaccharide chains (nonhydrolytic cleavage of glycosidic bonds). With the emergence of the high-throughput sequencing technologies, owing to the reduced cost and easier access, researchers have investigated the metabolic potential (referring to the gene content) of the microbial communities present in anaerobic environments. In a variety of studies, microbial communities of full-scale and lab-scale anaerobic digesters under optimal (10) and dysfunctioning conditions have been characterized (4, 11–13). Among bacteria, *Bacteroidetes* and *Firmicutes* have often been cited as the most dominant phyla, and their genomes were shown to contain high number of CAZyme-coding genes (14–16). CAZyme-coding genes can be expressed separately in bacterial genomes or they can form operons of coexpressed genes, e.g., cellulosomes in *Firmicutes* or polysaccharide-utilization loci (PULs) in *Bacteroidetes*. While cellulosomes have been given large scientific attention in the past (17, 18), PULs were much less investigated (17, 19). PULs are defined as gene clusters encoding cell envelope-associated enzymes that allow *Bacteroidetes* to bind to and degrade a specific polysaccharide and to import the released oligosaccharides inside the cell (20, 21). They are composed of at least one pair of *susC* and *susD* genes encoding a TonB-dependent transporter (TBDT) and a cell-

surface glycan-binding protein (SGBP), respectively (22). The protein SGBP is involved in the binding of carbohydrates while TBDT transports them through the outer membrane. The first PUL was identified from *Bacteroides thetaiotaomicron* isolated from the human gut and dedicated to starch degradation (23). During the last decades, a number of PULs have been discovered from omics studies, unravelling a wider diversity of CAZyme-coding genes than previously thought (24–28). One study reported the characterization of three PULs (two targeting xylan and one targeting pectins) using combined omics approaches, i.e., transcriptomics, proteomics and metabolomics, and originating from the marine environment (29). Another study identified a xyloglucan-specific PUL (from the human gut) through a metagenomics approach that was further experimentally validated (30).

The general purpose of the present study was to characterize the hydrolytic potential of the microbial community in an anaerobic digestion reactor. In particular, we aimed at assessing the carbohydrate hydrolytic potential in the context of an acidosis event, which is one common type of reactor dysfunction. Previous studies have shown that acidosis impacts microbial community structure at the taxonomic level. Here, we wanted to evaluate the community functional redundancy, with a special focus on the carbohydrate hydrolytic capacity of the anaerobic digester community. Metagenomic reconstruction allowed identification of 4,148 genes putatively assigned to CAZy families (including 49 putative PULs); only 1,052 of these were further assigned an Enzyme Commission (EC) class related to a potential hydrolytic function. To further link the community hydrolytic potential with key microbes, we investigated the newly reconstructed metagenome-assembled genomes (MAGs). Additional enzymatic activity assays were performed on a subset of predicted CAZymes belonging to a PUL named PUL219 from a dominant *Bacteroidetes* MAG in order to confirm their predicted activities.

RESULTS

Functional redundancy of hydrolytic metabolism under changing environmental conditions in AD reactors. In our previous study (4), the microbial composition of the community dynamics was investigated with the use of high-throughput 16S rRNA gene amplicon sequencing in triplicate lab-scale AD reactors operated for 300 days. Reactors were sequentially exposed to a decreasing hydraulic retention time (HRT) and increasing organic loading rate (OLR; phase I) of dry sugar beet pulp, leading to acidosis (phase II) and process recovery (phase III), as indicated in Fig. 1. Here, the hydrolytic potential of the microbial community from one of the reactors (named R3 in Goux et al. [4]) was further investigated with the use of metagenomics. Genomic DNA was extracted from seven samples taken at different stages of the experiment and sequenced using a high-throughput approach (31). As a result, over 30,000 contigs were reconstructed. Out of the 75,002 nonredundant protein coding genes identified in the reconstructed contigs, 4,148 (representing 5.53% of all coding sequences) were assigned to CAZy families (dbCAN analysis) (32), representing GH, GT, AA, CE, and PL, as well as CBM classes. In total, 1,052 putative CAZymes were further functionally classified into EC protein categories (Hotpep analysis) (33). At the microbiome level, we found a good correlation between the average metagenomic GH coding gene abundance at the different stages of the experiment and their gene copy numbers ($R^2 = 0.89$, $P < 0.001$) (Fig. 2A).

Selective pressure applied to the reactor led to a pH drop, and a significant variation over time in the abundance of specific microorganisms (bacterial MAGs) was observed (chi-square test P value, <0.001) (Fig. 1A to C). At the same time, the community gene profiles were conserved (based on the profile of the clusters of orthologous groups, or COGs [34]) (see Fig. S1A in the supplemental material), as well as acetogenesis and methanogenesis pathways (based on the profile of the Kyoto Encyclopedia of Genes and Genomes (KEGG) orthology [KO]) (35) (Fig. S1B and C), including during the acidosis stage and showing the community functional redundancy. The hydrolytic potential of the whole community (separately investigated based on the GH family profile and the

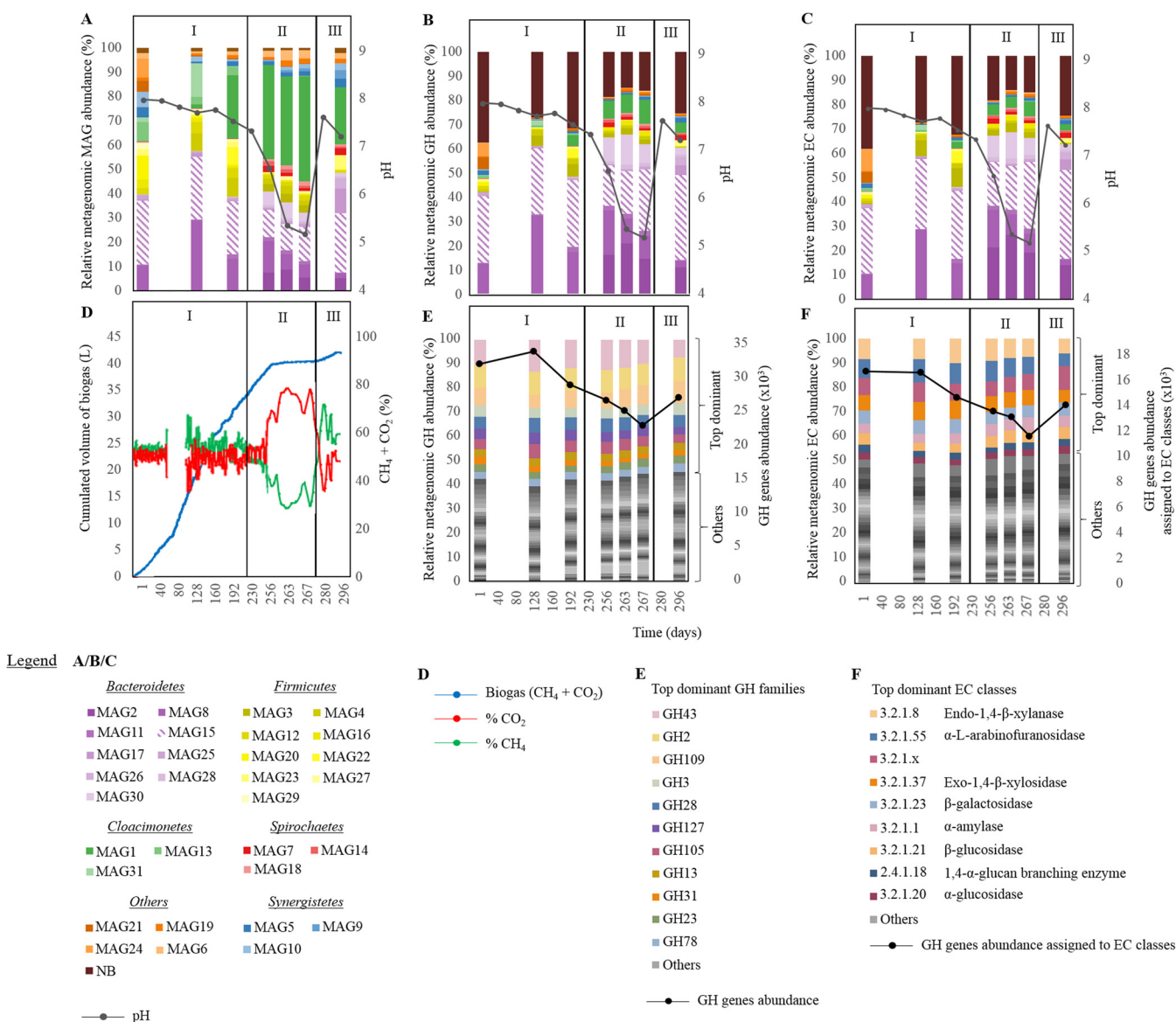


FIG 1 Metagenomics-assisted characterization of microbes and their hydrolytic potential throughout the anaerobic digestion (AD) experiment. (A) Relative metagenomic abundance of the different metagenome-assembled genomes (MAGs). (B) Relative metagenomic abundance of the different glycoside hydrolases (GH) colored according to the reconstructed MAG. (C) Relative metagenomic abundance of the different GHs additionally assigned to an Enzyme Commission (EC) category, colored according to the reconstructed MAG. (D) Characteristics of the biogas produced, including the cumulative biogas volume and the percentage of CH₄ and CO₂ in the biogas at the different stages of the experiment. (E) Relative metagenomic abundance of the different GHs colored according to the GH family. The top dominant families further discussed in the manuscript are highlighted. GH gene abundance, at the microbiome (whole-community) level, is represented by a black solid line. (F) Relative metagenomic abundance of the different GHs additionally assigned to an EC category, colored according to the EC category. The top dominant categories further discussed in the manuscript are highlighted. GH gene abundance assigned to EC classes is represented by a black solid line. Roman numerals indicate phases I to III of the AD experiment, as follows: I, decreasing hydraulic retention time (HRT) and increasing organic loading rate (OLR) of sugar beet pulp; II, acidosis; III, process recovery. NB, not binned.

EC profile) was also maintained during the whole AD experiment, despite a slight decrease of the abundance of genes coding for hydrolytic enzymes during the acidosis stage (Fig. 1E and F; Fig. S1D and E). Three GH families were overrepresented (based on their metagenomic abundances and numbers of the different genes present) (Fig. 1 and 2), including GH43 (average metagenomic abundance of $5.6\% \pm 1.1\%$ of the total CAZyme gene content), GH2 ($4.3\% \pm 0.3\%$), and GH109 ($3.4\% \pm 0.5\%$). Other families such as GH28 ($2.5\% \pm 0.3\%$), GH105 ($2.1\% \pm 0.4\%$), GH3 ($2.0\% \pm 2.0\%$), and GH13 ($2.0\% \pm 0.2\%$) were also well represented. Despite their identical CAZy family annotations, different hydrolytic activities (EC category) were assigned to proteins represent-

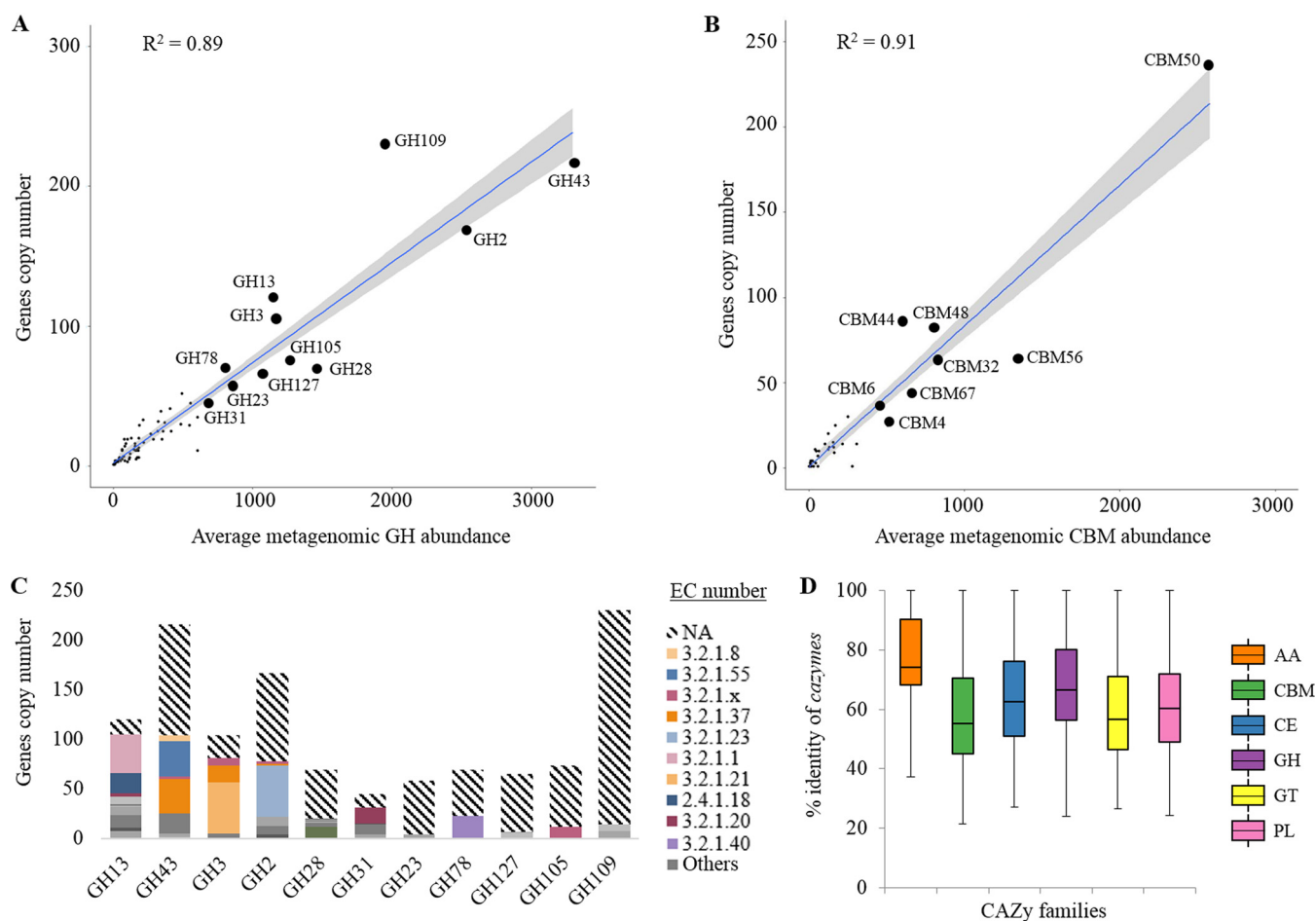


FIG 2 Characteristics of CAZy-coding genes identified in the anaerobic digestion (AD) reactor. (A) Correlation between the average metagenomic glycoside hydrolase (GH) abundance and the absolute number of gene copies at the microbiome (whole-community) level. Top dominant families further discussed in the manuscript are highlighted. (B) Correlation between the average metagenomic carbohydrate binding module (CBM) abundance and the absolute number of gene copies at the microbiome level. Top dominant CBM families are highlighted. (C) Assignment of the top dominant GHs to an Enzyme Commission (EC) category. NA, not assigned; EC 3.2.1.8, endo-1,4- β -xylanase; EC 3.2.1.55, α -L-arabinofuranosidase; EC 3.2.1.37, β -xylosidase; EC 3.2.1.23, β -galactosidase; EC 3.2.1.1, α -amylase; EC 3.2.1.21, β -glucosidase; EC 2.4.1.18, 1,4- α -glucan branching enzyme; EC 3.2.1.20, α -glucosidase; EC 3.2.1.40, α -L-rhamnosidase. (D) Box plot representations of the percentage identity of the CAZymes to proteins in the NCBI nonredundant protein database. The different CAZy families represented include auxiliary enzymes (AA), carbohydrate-binding modules (CBM), carbohydrate esterases (CE), glycoside hydrolases (GH), glycosyltransferases (GT), and polysaccharide lyases (PL).

ing a single CAZy family (Fig. 2C). Regarding their average metagenomic abundance, the hydrolytic portion of the metagenome was dominated by EC 3.2.1.8 (endo-1,4- β -xylanase; $8.2\% \pm 1.3\%$ of all CAZyme-coding gene assigned to an EC category), EC 3.2.1.55 (α -L-arabinofuranosidase; $7.7\% \pm 4.4\%$), EC 3.2.1.37 (exo-1,4- β -xylosidase; $6.5\% \pm 0.8\%$), and EC 3.2.1.23 (β -galactosidase; $5.3\% \pm 0.5\%$) (Fig. 1F). Similar to abundance of GHs, the average metagenomic abundance of CBM coding genes was highly correlated to the total gene copy numbers in the microbiome ($R^2 = 0.91$, $P < 0.001$) (Fig. 2B). The set of putative CBMs fell into 36 families (Fig. S1E), with CBM50 being the most abundant. In total, 91 coding genes from CBM50 were identified, and 20.9% of them were associated with a GH domain in a single protein.

Out of the identified CAZyme genes, 65.3% were further assigned to the different reconstructed MAGs. BLAST analysis against the NCBI nonredundant protein database showed overall low identities of the identified CAZymes for most of the MAGs (Fig. 2D). Roughly, 13.5% of predicted CAZymes showed high similarity ($\geq 95\%$) to the entries in the nonredundant protein database. This confirms the novelty of the novel enzymes. AA class seemed more conserved (higher sequence similarity), as previously demonstrated for the cow rumen metagenome (36). Interestingly, around 1.7% of all CAZyme-

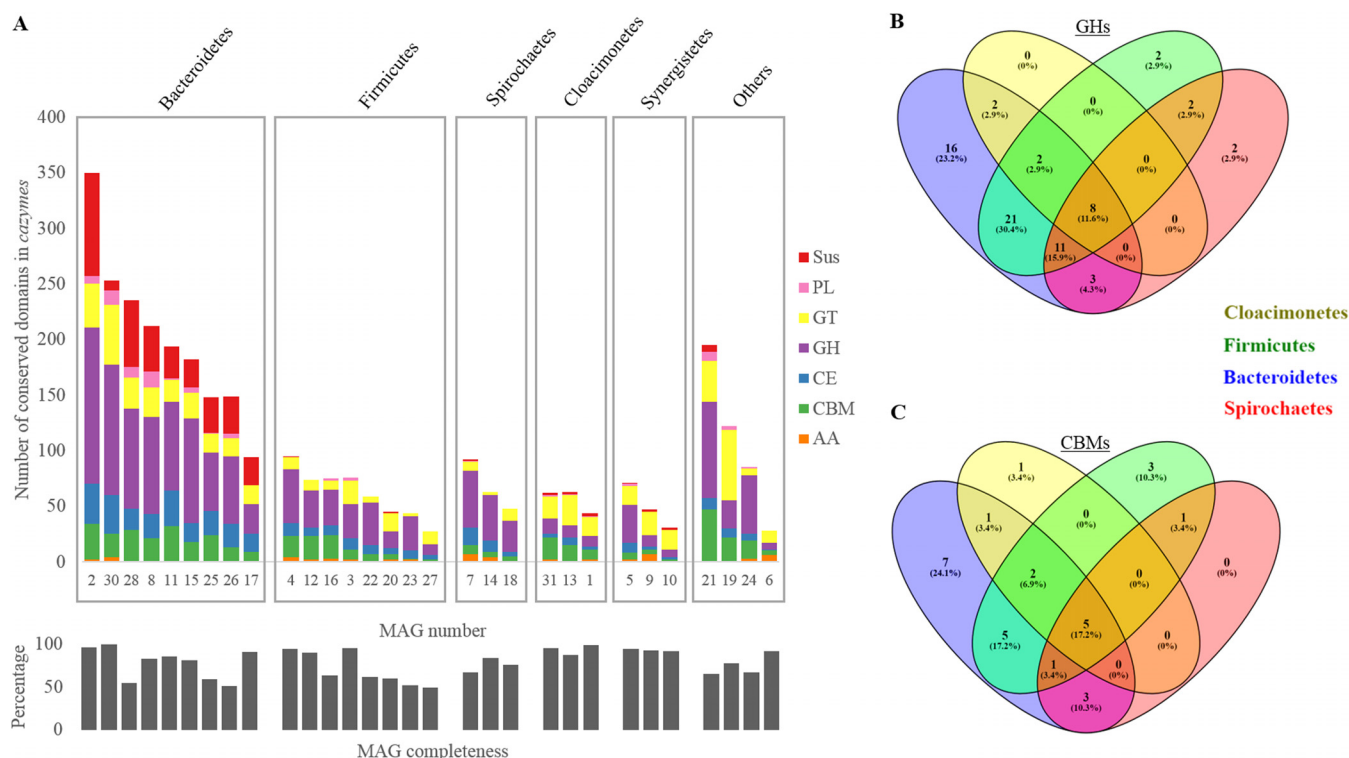
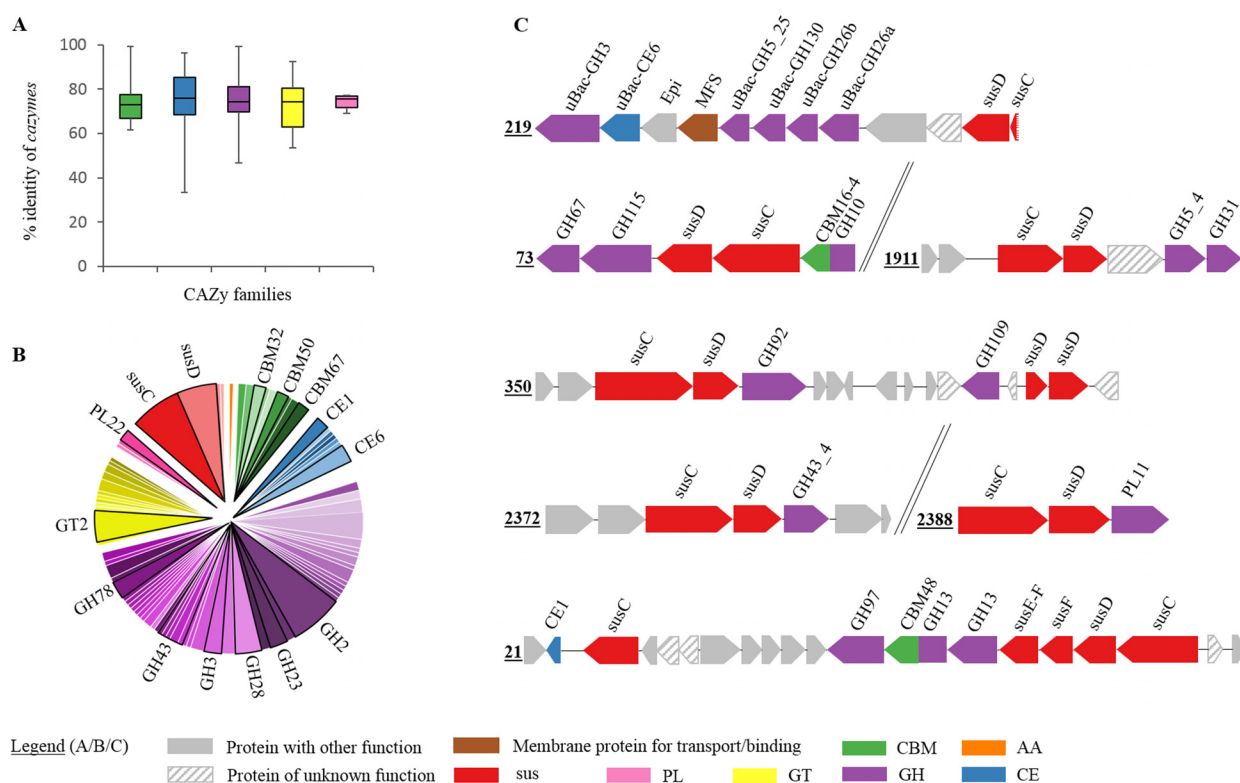


FIG 3 Characterization of the CAZy-coding genes (cazymes), including *sus*-like genes, in the reconstructed MAGs. (A) Distribution of the different conserved domains assigned to CAZy-coding genes in the reconstructed MAGs. (B and C) Venn diagram representation of the shared diversity of GHs and CBMs in MAGs assigned to the four main bacterial phyla present in the reactor. For the description of MAGs, refer to Table S1 in the supplemental material. The different CAZy families represented include auxiliary enzymes (AA), carbohydrate-binding modules (CBM), carbohydrate esterases (CE), glycoside hydrolases (GH), glycosyl-transferases (GT), and polysaccharide lyases (PL).

coding genes were annotated as AAs. The AA genomic content of *Spirochaetes* was much higher than that for the other MAGs ($5.4\% \pm 2.9\%$ of the CAZyme-coding genes versus $2.5\% \pm 1.9\%$ for *Firmicutes* and $0.6\% \pm 0.5\%$ for *Bacteroidetes*).

***Bacteroidetes* may be favored in some anaerobic digesters owing to its polysaccharide hydrolytic potential as reflected by higher CAZy diversity and genomic content.** We found that representatives of the phylum *Bacteroidetes* were among the most abundant MAGs across the digestion experiment (average metagenomic read abundance, $41.9\% \pm 8.4\%$), with MAG15 being dominant especially under optimal digestion conditions (Fig. 1A). Interestingly, nearly half (49.2%) of the binned CAZyme-coding genes were assigned to *Bacteroidetes* MAGs. On average, they accounted for $8.8\% \pm 1.8\%$ of their genome content, having higher frequency than the MAGs of *Firmicutes* ($3.6\% \pm 1.1\%$) or *Spirochaetes* ($2.9\% \pm 1.2\%$). Within *Bacteroidetes* MAGs, $50.1\% \pm 5.6\%$ of the CAZyme-coding genes were putative GHs, and $15.9\% \pm 4.0\%$ were putative CEs, showing the high carbohydrate hydrolytic potential of this phylum (Fig. 3A). The distribution into GHs and CEs was quite similar in *Firmicutes* ($49.9\% \pm 12.5\%$ and $12.6\% \pm 2\%$ of all CAZyme-coding genes) and slightly higher in *Spirochaetes* ($59.8\% \pm 4.7\%$ and $13.9\% \pm 4.9\%$). In contrast, reconstructed “*Candidatus* Cloacimonetes” and *Synergistetes* MAGs had higher contents of GTs than GHs ($39.7\% \pm 5.7\%$ versus $21.1\% \pm 2.75\%$ and $44.3\% \pm 19.0\%$ versus $31.6\% \pm 14.7\%$, respectively). Interestingly, partially reconstructed MAG21, assigned to *Planctomycetes* (65.5% genome completeness) (Table S1), harbored multiple CAZyme-coding genes (representing 8.8% of its total genomic content), mainly assigned to GH127, GH2, GH43, GH5, and CBM35 families.

The levels of diversity of assigned GH families and CBM families were the highest in the case of *Bacteroidetes* (Fig. 3B and C). Around 23.2% of GH families and 24.1% of CBM families were specific to *Bacteroidetes* MAGs only, while roughly 8.7% of GH families and



sented by GH2, GH43, GH28, GH78, and GH3 families (Fig. 4B), possibly allowing the species to cope with the different components of the sugar beet pulp. CBM67, CBM32, and CE6 domain-coding genes were also abundant in the MAG15 genome. Functionally assigned CAZyme-coding genes were represented by 31 EC classes with carbohydrate hydrolytic activity (Fig. S2D). Around 25 *sus*-like coding genes were identified in the genome of this new *Bacteroidetes* sp. and 76% of these genes were localized within PULs (Fig. 4C). In total, seven PULs were identified in MAG15. However, due to the novelty of CAZyme-coding genes, only a few could have been assigned to an EC class, therefore preventing the *in silico* substrate specificity prediction of the identified PULs. PUL219 was the most complete and contained diverse CAZyme-coding genes with putative functions predicted *in silico* to target acetylated glucomannan; therefore, it was further biochemically characterized.

Biochemical assays confirm predicted CAZy activities of six heterologously expressed proteins present in PUL219. PUL219 is composed of a putative pair of *susC* and *susD* homologs, one putative protein from the major facilitator superfamily (MFS), and six putative CAZymes belonging to five different families as well as a gene coding for a *N*-acyl-D-glucosamine 2-epimerase (Fig. 4C). Following dbCAN annotation, uBac-GH26a and -b were assigned to the GH26 family, while uBac-GH130, uBac-GH5, uBac-CE6, and uBac-GH3 were assigned to the GH130, GH5, CE6, and GH3 families, respectively. Following the Hotpep analysis, uBac-GH26a and -b were functionally assigned to EC 3.2.1.78 (β -mannanase), uBac-GH130 was assigned to EC 2.4.1.281 (4-O-beta-D-mannosyl-D-glucose phosphorylase), uBac-GH5 was assigned to EC 3.2.1.4 (endoglucanase), and uBac-GH3 was assigned to EC 3.2.1.37 (exo-1,4- β -xylosidase). These genes were successfully cloned, and the corresponding proteins were expressed in *Escherichia coli* Rosetta(DE3) cells.

Results of the enzymatic activity assays and *in silico* signal peptide prediction are summarized in Table S2 and Fig. 5A. As correctly predicted, uBac-GH26a and uBac-GH26b showed capability of endo-action on galactomannan and glucomannan (Table S2; Fig. 5A). They did not show any activity against the tested 4-nitrophenyl derivatives of monosaccharides, including 4-nitrophenyl α -D-mannopyranoside and 4-nitrophenyl β -D-mannopyranoside, showing no mannosidase activity with the tested substrates. Protein uBac-GH5 was active against 4-nitrophenyl β -D-cellobioside and was able to hydrolyze carboxymethyl cellulose (CMC). Moreover, it did not show any activity against 4-nitrophenyl β -D-glucopyranoside, confirming its predicted endoglucanase activity (EC 3.2.1.4). Finally, uBac-CE6 and uBac-GH3 were able to hydrolyze 4-nitrophenyl acetate and 4-nitrophenyl β -D-glucopyranoside, showing, respectively, esterase and β -glucosidase activities. The uBac-GH3 was shown to be a β -glucosidase (EC 3.2.1.21) rather than an exo-1,4- β -xylosidase (EC 3.2.1.37), as initially predicted by the Hotpep analysis. No activity was detected for uBAC-GH130 under the tested conditions and for the tested substrates. However, most of the characterized proteins assigned to this GH family exhibited phosphorylase activity (38). Additionally, the release of D-glucose, D-mannose, and acetic acid from acetylated konjac glucomannan was further assessed by the use of the different enzymatic cocktails and/or cascade reactions (Fig. 5A and Fig. S3). As expected, acetic acid was released when the substrate was treated with esterase (uBac-CE6); however, initial deacetylation did not increase subsequent release of D-glucose and D-mannose compared to the level of nondeacetylated konjac glucomannan. A larger amount of released D-glucose was measured when the enzymatic cocktail contained β -mannanase (cytoplasmic uBac-GH26a) and β -glucosidase (uBac-GH3). For a comparison, a smaller amount of released D-glucose residues was detected when the putatively periplasmic β -mannanase (uBac-GH26b) was combined in a cocktail with β -glucosidase. This finding suggests differential specificity of the two β -mannanases and indicates the formation of, respectively, longer and shorter mannooligosaccharides when the periplasmic and cytoplasmic isoforms are used, respectively. Similarly, a larger amount of released D-mannose was observed when the enzymatic cocktail contained the cytoplasmic form of the β -mannanase combined with the β -glucosidase (uBac-GH3) and independently of the prior substrate

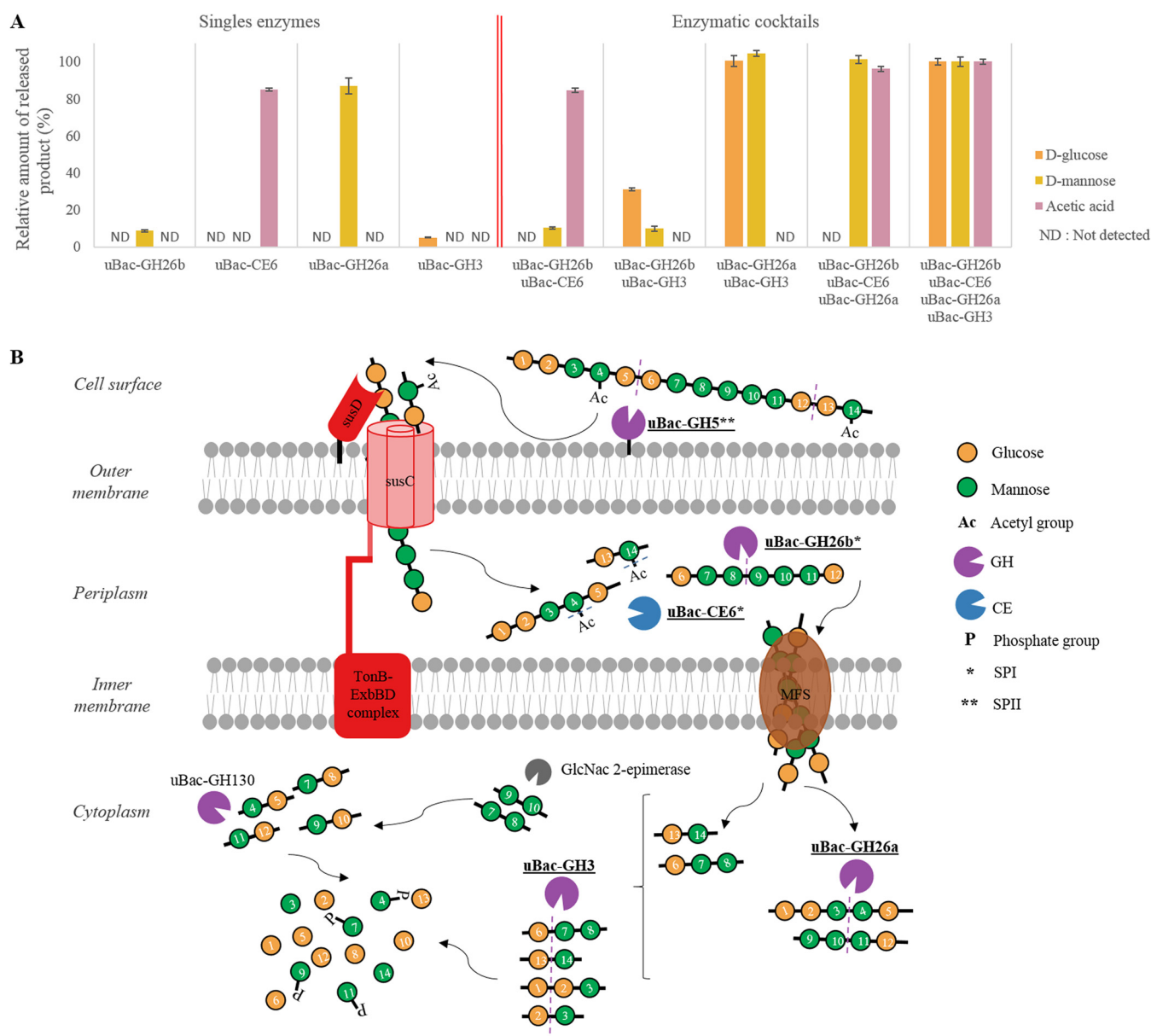


FIG 5 Characterization of PUL219 toward acetylated glucomannan degradation. (A) Release of D-glucose, D-mannose, and acetic acid after enzymatic hydrolysis of acetylated konjac glucomannan. Single enzymes as well as enzymatic cocktails were tested; for all the reactions, enzymatic hydrolysis of substrate lasted 1 h at 37°C. An enzymatic cocktail containing uBac-GH3, uBac-CE6, uBac-GH26a, and uBac-GH26b was taken as a reference to normalize results to 100%. (B) Proposed mechanism of action of the putative acetylated glucomannan-targeting polysaccharide utilization locus PUL219, isolated from a new *Bacteroidetes* species. GlcNac 2-epimerase, *N*-acetyl-D-glucosamine 2-epimerase; SPI, signal peptidase I predicted by LipoP, version 1.0; SPII, signal peptidase II predicted by LipoP, version 1.0. Enzymes biochemically characterized in this study are highlighted in bold and underlined.

deacetylation (Fig. S3). Additionally, using uBac-GH26a as a single enzyme resulted in mannose released from the treated konjac glucomannan. This result in combination with preliminary activity tests with synthetic substrates (e.g., no mannosidase activity detected using 4-nitrophenyl β -D-mannopyranoside) suggests that small-chain oligosaccharides from treated konjac glucomannan could be hydrolyzed by uBac-GH26a, releasing mannoses and shortened manno-oligosaccharides.

DISCUSSION

Functional redundancy of AD microbiome. On one hand, hydrolysis is recognized as a bottleneck of the AD process, especially in the case of highly lignified substrates (8). On the other hand, an increased OLR of easily digestible substrates promotes fast

hydrolysis, leading to VFA accumulation and the resulting process failure, known as acidosis. In a previous study, we showed that acidosis, promoted by increasing OLR of dry sugar beet pulp, affected the microbial community structure, causing decreased methane production (Fig. 1A to D) (4). During the acidosis phase, some CO₂ was still produced at a low rate compared to that under optimal conditions (linked to a lower total number of CAZyme-coding genes observed during acidosis in our study) (Fig. 1), confirming the activity of hydrolytic bacteria at lower pH (39). Following the reestablishment of the optimal AD conditions in the reactor, increased hydrolysis was detected (measured by increased amount of the produced biogas) (Fig. 1), suggesting that the hydrolytic readiness of the microbial community was maintained during the process failure. In addition, we intended to investigate the microbiome functional redundancy with a special focus on its carbohydrate hydrolytic capacity, and to that purpose we applied metagenomics to seven samples taken at different stages of the experiment. Metagenomic results confirmed a significant variation in the abundances of individual bacteria over the experiment, based on the newly reconstructed MAGs (Fig. 1). This is consistent with the previously performed 16S rRNA gene amplicon high-throughput characterization of the microbial community (4). Functional redundancy, known to be widespread in different microbial environments (7), was previously shown at the metaproteome level for a steady-state anaerobic reactor (40). In our study, we showed that despite the differential genomic content of individual MAGs, a strong community functional redundancy was retained even during severe acidosis. This could explain why the reactor was able to quickly recover from acidosis and to promptly restart the production of biogas (functional profiles of genes involved in acetogenic and methanogenic pathways were largely retained as well) (see Fig. S1B and C in the supplemental material). Taxonomic assignment of CAZyme-coding genes showed a shift in the microbial population capable to hydrolyze the substrate along the experiment (Fig. 1B and C). Nevertheless, the genetic potential highlighted at the community-wide level by nearly identical CAZyme-coding gene profiles remained unchanged. This observation somehow fits the recently proposed premise in the context of the human gut of “function first, taxa second” (41). At the same time at the gene expression level, higher variability and sensitivity to perturbation have previously been observed than indicated by the content of the metagenome. Therefore, further metatranscriptomics study would complement the proposed functional redundancy of the AD microbiome by assessing its functional plasticity and, thus, its ability to adapt to perturbations by modulating the expression of the different genes.

Diversity of CAZymes. CAZymes are very important to the success of microbes in AD reactors fed with vegetal substrate as they are involved in the digestion of complex polysaccharides, which are particularly abundant in this environment (17). Here, we showed the usefulness of the complementary Hotpep analysis to the dbCAN-mediated CAZyme-coding gene discovery and classification. Indeed, within the abundant GH families (e.g., GH2, GH43, and GH3) (Fig. 1E) different hydrolytic activities were detected on the basis of the assigned EC classes, e.g., α -L-arabinofuranosidase (EC 3.2.1.55), exo-1,4- β -xylosidase (EC 3.2.1.37), endo-1,4-beta-xylanase (EC 3.2.1.8), β -glucosidase (EC 3.2.1.21), and β -galactosidase (EC 3.2.1.23) (Fig. 2C). Still, the biochemical characterization of each new enzyme was necessary to provide correct insights into its saccharolytic capacity and substrate specificity. Sugar beet pulp is composed of diverse carbohydrates, including pectins enriched in arabinans and galactans, cellulose, and hemicelluloses, which comprise small amounts of glucomannans and xyloglucans (42). Accordingly, CAZymes assigned to the most abundant EC categories were predicted to be mainly involved in the hydrolysis of hemicellulosic material and especially (arabino)xylan, (arabino)galactan, and xyloglucan-enriched hemicelluloses. Thus, the ability of the microbiome to hydrolyze sugar beet pulp could be further linked to the abundance of putative β -galactosidases (EC 3.2.1.23) as well as α -L-arabinofuranosidases (EC 3.2.1.55), which were shown to be largely involved in galactan and arabinan degradation (43, 44). Sugar beet pulp constituent-targeting

CBMs were also abundant, including pectin-targeting CBM32 or CBM67 that were shown to bind to polygalacturonic acid or α -L-rhamnoside, respectively (45, 46).

The capacity of the microbiome to use different carbon sources was also confirmed by the presence of other CBMs, with CBM50 and CBM56 targeting chitin and peptidoglycan (e.g., fungal and bacterial cell wall components, respectively) being the most abundant. Moreover, family GH109 has often been recorded in metagenomic studies of anaerobic digesters (47), including in our reactor. This family is known to express the α -N-acetylgalactosaminidase activity (EC 3.2.1.49), possibly targeting acetylgalactosamine, which is a component of bacterial cell walls (37). On one hand, the abundance of genes assigned to this family, together with the presence of CBM50 and CBM56 domains, might be linked to the expected high turnover rates of bacterial biomass resulting from high competition in this environment (48). On the other hand, only a few GH109-classified CAZyme-coding genes were assigned to the respective EC 3.2.1.49 category ($0.4\% \pm 0.3\%$ of metagenomic abundance) (Fig. 2C). Therefore, further investigation is required to confirm the specific (hydrolytic) activity of this family of proteins. The presence of AA coding genes, including AA2-containing class II lignin-modifying peroxidases, indicates the potential to decompose lignin by some anaerobic digester microbes. Demethylation of lignin was suggested for some *Spirochaetes* isolated from the termite gut (49), which is in accordance to our *Spirochaetes* MAGs being enriched in AA coding genes.

Bacteroidetes and their PULs. Similar to findings of previous studies of AD microbiomes (10, 50, 51), *Bacteroidetes* MAGs were shown to be among of the most dominant in the reconstructed metagenome representing the digester microbial community. Indeed, higher genomic content and functional diversity of hydrolytic genes (GHs and ECs) in *Bacteroidetes* MAGs (Fig. 3A and Fig. S4) could ensure digestibility of a wide range of substrates, thus favoring their abundance in AD reactors under certain environmental conditions (10). Individual *Bacteroidetes* MAGs were characterized with distinct CAZyme-coding gene functional profiles suggesting niche differentiation (different nutritional requirements) of the different species (Fig. S2). However, at the phylum level the *Bacteroidetes* population, by regulating the abundance of single individuals, was capable of ensuring the stability of gene functional profiles, thus largely contributing to the community-wide functional redundancy. *Bacteroidetes* and some other bacteria (e.g., *Firmicutes* and representatives of *Spirochaetes*, *Lentisphaerae*, *Planctomycetes*, and *Thermotogae*) that harbor multiple CAZyme-coding genes in their genomes (Table S1) seem responsible for complex carbohydrate degradation (14). Interestingly, *Planctomycetes* MAG21 harbors a CAZy profile similar to that of *Bacteroidetes* MAGs with a relatively high CAZyme-coding gene content (Fig. 3A). Even though the role of this bacterium in anaerobic digestion has not been widely discussed, previous studies suggest the involvement of *Planctomycetes* in degradation of chitin and cellulose in agricultural soil (52) and humus decomposition in the termite gut (53). These observations might indicate a similar hydrolytic potential to *Bacteroidetes* and the importance of *Planctomycetes* in the anaerobic digestion process (due to its ability to decompose recalcitrant organic matter). In contrast, “*Ca. Cloacimonetes*” and *Synergistetes* have different CAZy profiles, mainly enriched in GTs, and seem of less importance for carbohydrate hydrolysis. Nevertheless, secondary ion mass spectrometry-*in situ* hybridization using an iodine-labeled oligonucleotide probe combined with high-resolution nanometer-scale secondary-ion mass spectrometry (SIMS) supported the role of “*Ca. Cloacimonetes*” in the fermentative digestion of cellulose (54). Owing to its status of a candidate phylum and to the recently proposed diversity of “*Ca. Cloacimonetes*” in the different full-scale AD reactors (10), a broad range of functional diversity may be expected within this phylum.

MAG15 assigned to an unknown *Bacteroidales* was characterized with relatively high community abundance in the reactor during the whole AD experiment (Fig. 1A). According to its CAZy profile (Fig. 4B and C), it showed potential to digest a broad

range of polysaccharides present in sugar beet pulp. CAZyme-coding genes assigned to the family GH78 known to have α -L rhamnosidase activity (EC 3.2.1.40) acting on rhamnogalacturonan, a component of pectins (44), were present in multiple copies in its genome. In addition, the presence of numerous CAZyme-coding gene modules from the CBM67 family having α -L-rhamnose binding activity (46) confirms the potential of this bacterium to target pectins (Fig. 4B). The presence of (hetero)xylan targeting CAZyme genes, including the putative enzymes attacking the backbone (endoxylanases and xylosidases assigned to families GH2, -3, -10, and -43, e.g.), as well as arabinases (e.g., GH2 and 43) and galactosidases (e.g., GH2), confirms the ability of MAG15 to decompose different forms of xylans. Many novel *Bacteroidetes* were shown to contain various sets of PULs predicted to be involved in complex hydrolysis of diverse carbohydrates (19), changing our understanding of polysaccharide metabolism inside this phylum. The partially reconstructed MAG15 contains multiple PULs, including PUL73 predicted to target xylan and starch-targeting PUL21. PUL219, containing the highest diversity of putative CAZyme-coding genes, was additionally biochemically validated to target (acetylated) glucomannan. Glucomannan is a polysaccharide composed of β -D-1,4-linked mannose and glucose monomers (possible combinations include glucose-mannose, glucose-glucose, mannose-glucose, and mannose-mannose linkages), with side chain acetyl groups (Fig. 5B) (55, 56). Previously biochemically validated PULs included a galactomannan-targeting PUL from a marine *Bacteroidetes* (57) and a cellulose-degrading one from a rumen-isolated bacterium (58). Here, we propose a model of PUL-assisted acetylated glucomannan degradation for a *Bacteroidetes* species (Fig. 5B). This model was reconstructed based on the activity tests performed in this study using purified proteins and complementary *in silico* predictions, as described below. According to the predicted presence of a lipoprotein signal peptidase II and a cleavage site within uBac-GH5 sequence, this endoglucanase is most probably attached to the outer membrane, extracellularly digesting acetylated glucomannans and releasing acetylated glucomannan oligosaccharides (59, 60). As *susD* is attached to the outer membrane, it binds the resulting substrate and directs it to the *susC* transporter, which passes it to the periplasm (61). A signal peptidase I and a cleavage site were *in silico* predicted in both uBac-GH26b and uBac-CE6, suggesting their migration from the cytoplasm to the periplasm (62). In the periplasm, β -mannanase (uBac-GH26b) further hydrolyzes the acetylated glucomannan oligomers while acetyl groups are removed by the action of an esterase (uBac-CE6). Even shorter oligosaccharides may now be transported through the inner membrane with the help of the MFS (63). In the cytoplasm, short chains of glucomannan are further digested by uBac-GH26a (second β -mannanase), possibly releasing mono-, di-, and trisaccharides (64). These oligosaccharides are further hydrolyzed by uBac-GH3 (β -glucosidase), releasing glucose, mannose, and mannobiose. *N*-Acyl-D-glucosamine 2-epimerase transforms mannobiose into β -D-mannosyl-(1 \rightarrow 4)-D-glucose (MannosylGlu). Hydrolysis of MannosylGlu to glucose and mannose-1-phosphate is mediated by uBac-GH130 (presumed to be the 4-O-beta-D-mannosyl-D-glucose phosphorylase) (65, 66). A similar galactomannan-targeting PUL from *Bacteroides ovatus* has been previously characterized (67), including two β -mannanases from the GH26 family (BoMan26A and BoMan26B). In that study, BoMan26B was shown to release longer-chain oligosaccharides than BoMan26A, similarly to uBac-GH26b and uBac-GH26a as proposed in our model (Fig. 5). Phylogenetic analysis of several mannanases (including previously characterized BoMan26A and BoMan26B) confirmed the presence of two distinct mannanase clusters inside the GH26 family, showing the differential clustering of uBac-GH26a and uBac-GH26b, which are closely related to BoMan26A and BoMan26B, respectively (Fig. S5).

Further characterization of novel PULs should bring even more interesting insights into the complex carbohydrate metabolism in *Bacteroidetes*. Specific substrate targeting ready-to-use enzymatic cocktails naturally selected in PULs might be of high interest to the biotechnology sector as well, especially for developing biomass-based green chemistry. For example, acetylated glucomannan is the main hemicellulosic

component of the secondary cell walls in soft woods (68), which are common substrates for pulp and paper production. A nature-evolved cocktail of acetylated glucomannan-targeting enzymes, as characterized in, e.g., PUL219, might be an interesting alternative to the currently used chemical pretreatments applied in these industries.

MATERIALS AND METHODS

Sampling, metagenomics, and data processing. Samples were taken from an R3 lab-scale continuously stirred tank reactor (CSTR) of 100-liter capacity fed with dry sugar beet pulp as described in a previous study. Decreasing HRT and increasing OLR resulted in a pH decrease and lower biomass-to-methane conversion ratio (4). In total, seven time points were analyzed, corresponding to the different stages of the experiment (Fig. 1). Total genomic DNA was extracted with a PowerSoil DNA isolation kit (MoBio Laboratories, Inc.), according to the manufacturer's instructions. Metagenomic DNA libraries were constructed, sequenced, and analyzed as previously described (31). Binning of assembled contigs resulted in over 30 MAGs with average genome completeness of $76.6\% \pm 19.8\%$ and contamination below 5%. MAGs described in this study correspond to the bacterial bins listed in Table 2 in a previous publication (31). The taxonomic assignment of MAGs was done with PhyloPhlan (69). Prodigal was used to predict coding sequences on the coassembled contigs (70). CAZyme-coding genes were searched with the dbCAN (dbCAN-fam-HMMs.txt.v6 [27]) against a CAZy database (<http://www.cazy.org/>). The resulting CAZyme-coding genes were manually curated. A difference in abundances of MAGs/GHs/ECs was tested by a chi-square test. PULs were predicted according to the PUL database (71), and the presence of at least one *susD* or *susC* gene was mandatory for partial PUL prediction. Homology to peptide pattern (Hotpep) was used to assign an EC class to the identified CAZymes (72). Taxonomic and functional annotation of the coassembled contigs was done with IMG-MER (73) and corresponds to the project identification number 3300002079. The most complete MAGs were reannotated with RAST (79).

CAZyme-coding gene heterologous expression in *E. coli*. Details for gene selection, isolation, and cloning are shown in Text S1 in the supplemental material. For CAZyme-coding gene expression, plasmid pET-52b(+) (Millipore Corporation, Billerica, MA, USA) and the *E. coli* Rosetta(DE3) strain (Millipore Corporation, Billerica, MA, USA) were used. Genes were ligated into pET-52b(+) vector and introduced in *E. coli* by heat shock. An isolated colony (the insert was determined by sequencing) was grown in LB medium containing ampicillin (LB-Amp) overnight at 37°C with 250 rpm shaking. In total, 100 μ l of this culture was transferred into a fresh 100 ml of LB-Amp. Cells were grown at 37°C with shaking at 250 rpm to an optical density at 600 nm (OD_{600}) of 0.5, and isopropyl- β -D-thiogalactopyranoside (IPTG; Thermo-Fisher, Waltham, MA, USA) was added at final concentration of 0.5 mM to induce the expression of recombinant proteins. The culture was maintained for 20 h at room temperature and 250 rpm.

Cells were collected by centrifugation at $5,000 \times g$ at 4°C for 15 min. Supernatant was stored at -20°C , and the cell pellet was redissolved in an appropriate lysis buffer (50 mM NaH_2PO_4 , 300 mM NaCl, 10 mM imidazole, pH 8). As a negative control, an empty pET-52b(+) vector was cloned, expressed, and processed as a recombinant protein sample.

Using a method adapted from previous studies (74, 75), cell lysis was achieved using a Sonicator VC750 (Sonics & Materials, Inc., Newtown, CT, USA) with a 2-min pulse (1 s on/1 s off) followed by a 2-min pause (40% of amplitude). This step of sonication was repeated twice. Samples were transferred into 2-ml tubes and centrifuged for 15 min at $16,000 \times g$ at 4°C. The liquid part was filtered using an Acrodisc 13-mm syringe filter with a 0.2- μ m-pore-size Supor membrane (Millipore Corporation, Billerica, MA, USA). Protein detection and partial purification were performed using a histidine tag located at the C terminus. Initial detection on cell lysates was performed using Western blotting. Separation was done using 10% Mini-Protean TGX precast protein gels (Bio-Rad, Hercules, CA, USA). Transfer was performed with a Trans-Blot Turbo transfer system (Bio-Rad, Hercules, CA, USA). Detection was carried out using an appropriate antibody labeled with horseradish peroxidase (HRP) (6 \times His tag polyclonal antibody) (ThermoFisher, Waltham, MA, USA). Ni-nitrilotriacetic acid (NTA) agarose matrix was used to load polypropylene columns for partial purification according to the supplier's instructions (Qiagen, Hilden, Germany). Protein quantification was determined using a reducing agent and detergent-compatible (RC DC) protein assay (Bio-Rad, Hercules, USA) and bovine serum albumin as a standard. Proteins were identified, and details can be found in Text S2 in the supplemental material.

Signal peptide prediction and activity assays. Signal peptides were predicted using LipoP, version 1.0 (<http://www.cbs.dtu.dk/services/LipoP/>) (76). Enzymatic assays were performed using 4-nitrophenyl derivatives as substrates (Sigma-Aldrich, Saint Louis, MO, USA). Initially, 50 μ l of partially purified protein solution was incubated with 100 μ l of citrate phosphate buffer, pH 7 (0.1 M citric acid, 0.2 M dibasic sodium phosphate), and 50 μ l of substrate (Table S3). A microplate was incubated at 37°C in a Tecan Spark 20 M microplate reader (Tecan, Männedorf, Switzerland). The rate of released 4-nitrophenol was monitored at 405 nm. Assays were performed in triplicate.

Following previous studies (77, 78), enzymatic activity was also assessed by the release of reducing sugars. CMC (Sigma-Aldrich, Saint Louis, USA), arabinoxylan, galactomannan, glucomannan, and xylan (Megazyme, Wicklow, Ireland) were used as substrates. Briefly, 100 μ l of partially purified enzyme solution was incubated with 50 μ l of citrate phosphate buffer, pH 7 (0.1 M citric acid, 0.2 M dibasic sodium phosphate), and 100 μ l of substrate at 37°C for 30 min (Table S3). The concentrations of reducing sugars were determined applying the Somogyi-Nelson method, and absorbance was read at 620 nm using Specord Plus spectrophotometer (Analytik Jena, Jena, Germany). Assays were performed in triplicates.

Single enzymes as well as enzymatic cocktails were used for the pretreatment of glucomannan to determine the release of D-glucose, D-mannose, and acetic acid. Briefly, pretreatment was carried out in 500 μ l as follows: 250 μ l of substrate, 50 μ l of buffer, and 50 μ l of each enzyme assessed; final volume was reached by addition of water if necessary. The reaction mixture was incubated at 37°C for 1 h. If a cascade reaction was tested, a second enzyme (enzymatic set) was added after 1 h of pretreatment and reincubated for 1 h at 37°C. The release of D-glucose, D-mannose, and acetic acid was determined using commercially available kits (Megazyme, Wicklow, Ireland), according to the supplier's instructions. An enzymatic cocktail containing uBac-GH3, uBac-CE6, uBac-GH26a, and uBac-GH26b was taken as a reference to normalize results to 100%.

Data availability. The most complete MAGs can be accessed using a RAST (<http://rast.nmpdr.org>) guest account under the following identification numbers: 6666666.364478 (MAG1), 6666666.364479 (MAG2), 6666666.364480 (MAG3), 6666666.364481 (MAG4), 6666666.364482 (MAG5), 6666666.364483 (MAG6), 6666666.364484 (MAG7), 6666666.364485 (MAG8), 6666666.364486 (MAG9), 6666666.364487 (MAG10), 6666666.364489 (MAG11), 6666666.364490 (MAG12), 6666666.364491 (MAG13), 6666666.364493 (MAG14), 6666666.364494 (MAG15), 6666666.364495 (MAG16), 6666666.364496 (MAG17), 6666666.364497 (MAG18), 6666666.364498 (MAG19), 6666666.364499 (MAG20), 6666666.364500 (MAG21), 6666666.364502 (MAG23), 6666666.364503 (MAG25), 6666666.364504 (MAG30), and 6666666.364505 (MAG31).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/AEM.00895-19>.

SUPPLEMENTAL FILE 1, PDF file, 1.6 MB.

ACKNOWLEDGMENTS

We thank Dominika Klimek, Sebastien Lemaigre, and Stephanie Giusti-Miller for their technical support and Kjell Sergent for his expertise in protein identification analysis.

This work was supported by the Luxembourg National Research Fund, as follows: by FNR CORE 2011 project GASPOP (C11/SR/1280949: Influence of the Reactor Design and the Operational Parameters on the Dynamics of the Microbial Consortia Involved in the Biomethanation Process) and by FNR CORE 2014 project OPTILYS (C14/SR/8286517: Exploring the higher termite lignocellulolytic system to optimize the conversion of biomass into energy and useful platform molecules).

REFERENCES

- Pain BF, Hephherd RQ. 1985. Anaerobic digestion of livestock wastes, p 9–14. In Pain BF, Hephherd RQ (ed), *Anaerobic digestion of farm waste*. NIRD technical bulletins. NIRD, Reading, United Kingdom.
- Ahring BK. 2003. Perspectives for anaerobic digestion. *Adv Biochem Eng Biotechnol* 81:1–30.
- Lemaigre S, Adam G, Goux X, Noo A, De Vos B, Gerin PA, Delfosse P. 2016. Transfer of a static PCA-MSPC model from a steady-state anaerobic reactor to an independent anaerobic reactor exposed to organic overload. *Chemom Intell Lab Syst* 159:20–30. <https://doi.org/10.1016/j.chemolab.2016.09.010>.
- Goux X, Calusinska M, Lemaigre S, Marynowska M, Klocke M, Udelhoven T, Benizri E, Delfosse P. 2015. Microbial community dynamics in replicate anaerobic digesters exposed sequentially to increasing organic loading rate, acidosis, and process recovery. *Biotechnol Biofuels* 8:122. <https://doi.org/10.1186/s13068-015-0309-9>.
- Gerardi MH. 2003. *The microbiology of anaerobic digesters*, 1st ed. John Wiley & Sons, Inc., Hoboken, New Jersey.
- Lerm S, Kleyböcker A, Miethling-Graff R, Alawi M, Kasina M, Liebrich M, Würdemann H. 2012. Archaeal community composition affects the function of anaerobic co-digesters in response to organic overload. *Waste Manag* 32:389–399. <https://doi.org/10.1016/j.wasman.2011.11.013>.
- Louca S, Polz MF, Mazel F, Albright MBN, Huber JA, O'Connor MI, Ackermann M, Hahn AS, Srivastava DS, Crowe SA, Doebeli M, Parfrey LW. 2018. Function and functional redundancy in microbial systems. *Nat Ecol Evol* 2:936–943. <https://doi.org/10.1038/s41559-018-0519-1>.
- Álvarez C, Reyes-Sosa FM, Díez B. 2016. Enzymatic hydrolysis of biomass from wood. *Microb Biotechnol* 9:149–156. <https://doi.org/10.1111/1751-7915.12346>.
- Coutinho P, Henrissat B. 1999. Carbohydrate-active enzymes: an integrated database approach, p 3–12. In Gilbert H, Davies G, Henrissat H, Svensson B (ed), *Recent advances in carbohydrate bioengineering*. The Royal Society of Chemistry, Cambridge, United Kingdom.
- Calusinska M, Goux X, Fossépré M, Muller EEL, Wilmes P, Delfosse P. 2018. A year of monitoring 20 mesophilic full - scale bioreactors reveals the existence of stable but different core microbiomes in bio - waste and wastewater anaerobic digestion systems. *Biotechnol Biofuels* 11:196. <https://doi.org/10.1186/s13068-018-1195-8>.
- Sun L, Pope PB, Eijssink VGH, Schnürer A. 2015. Characterization of microbial community structure during continuous anaerobic digestion of straw and cow manure. *Microb Biotechnol* 8:815–827. <https://doi.org/10.1111/1751-7915.12298>.
- Goux X, Calusinska M, Fossépré M, Benizri E, Delfosse P. 2016. Start-up phase of an anaerobic full-scale farm reactor - Appearance of mesophilic anaerobic conditions and establishment of the methanogenic microbial community. *Bioresour Technol* 212:217–226. <https://doi.org/10.1016/j.biortech.2016.04.040>.
- Lemaigre S, Adam G, Gerin PA, Noo A, De Vos B, Klimek D, Goux X, Calusinska M, Delfosse P. 2018. Potential of multivariate statistical process monitoring based on the biogas composition to detect free ammonia intoxication in anaerobic reactors. *Biochem Eng J* 140:17–28. <https://doi.org/10.1016/j.bej.2018.08.018>.
- Campanaro S, Treu L, Kougias PG, Luo G, Angelidaki I. 2018. Meta-genomic binning reveals the functional roles of core abundant micro-organisms in twelve full-scale biogas plants. *Water Res* 140:123–134. <https://doi.org/10.1016/j.watres.2018.04.043>.
- Svartström O, Alneberg J, Terrapon N, Lombard V, De Bruijn I, Malmsten J, Dalin AM, El Muller E, Shah P, Wilmes P, Henrissat B, Aspeborg H, Andersson AF. 2017. Ninety-nine de novo assembled genomes from the moose (*Alces alces*) rumen microbiome provide new insights into microbial plant biomass degradation. *ISME J* 11:2538–2551. <https://doi.org/10.1038/ismej.2017.108>.
- Vanwonterghem I, Jensen PD, Rabaey K, Tyson GW. 2016. Genome-centric resolution of microbial diversity, metabolism and interactions in anaerobic digestion. *Environ Microbiol* 18:3144–3158. <https://doi.org/10.1111/1462-2920.13382>.
- Kougias PG, Campanaro S, Treu L, Tsapekos P, Armani A, Angelidaki I.

2018. Spatial distribution and diverse metabolic functions of lignocellulose-degrading uncultured bacteria as revealed by genome-centric metagenomics. *Appl Environ Microbiol* 84:1–14. <https://doi.org/10.1128/AEM.01244-18>.
18. Gefen G, Anbar M, Morag E, Lamed R, Bayer E. a. 2012. Enhanced cellulose degradation by targeted integration of a cohesin-fused -glucosidase into the *Clostridium thermocellum* cellulosome. *Proc Natl Acad Sci U S A* 109:10298–10303. <https://doi.org/10.1073/pnas.1202747109>.
19. Grondin JM, Tamura K, Déjean G, Abbott DW, Brumer H. 2017. Polysaccharide utilization loci: fuelling microbial communities. *J Bacteriol* 199:e00860-16. <https://doi.org/10.1128/JB.00860-16>.
20. Bjursell MK, Martens EC, Gordon JL. 2006. Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, *Bacteroides thetaiotaomicron*, to the suckling period. *J Biol Chem* 281:36269–36279. <https://doi.org/10.1074/jbc.M606509200>.
21. Terrapon N, Lombard V, Gilbert HJ, Henrissat B. 2015. Automatic prediction of polysaccharide utilization loci in *Bacteroidetes* species. *Bioinformatics* 31:647–655. <https://doi.org/10.1093/bioinformatics/btu716>.
22. Martens EC, Koropatkin NM, Smith TJ, Gordon JL. 2009. Complex glycan catabolism by the human gut microbiota: The bacteroidetes sus-like paradigm. *J Biol Chem* 284:24673–24677. <https://doi.org/10.1074/jbc.R109.022848>.
23. Anderson KL, Salyers AA. 1989. Biochemical evidence that starch breakdown by *Bacteroides thetaiotaomicron* involves outer membrane starch-binding sites and periplasmic starch-degrading enzymes. *J Bacteriol* 231:3192–3198. <https://doi.org/10.1128/jb.171.6.3192-3198.1989>.
24. Rogowski A, Briggs JA, Mortimer JC, Tryfona T, Terrapon N, Morland C, Day AM, Zheng H, Rogers TE, Lowe EC, Basle A, Thompson P, Hawkins AR, Yadav MP, Henrissat B, Martens EC, Dupree P, Gilbert HJ, Bolam DN. 2015. Glycan complexity dictates microbial resource allocation in the large intestine. *Nat Commun* 6:7481. <https://doi.org/10.1038/ncomms8481>.
25. Zhang M, Chekan JR, Dodd D, Hong P-Y, Radlinski L, Revindran V, Nair SK, Mackie RI, Cann I. 2014. Xylan utilization in human gut commensal bacteria is orchestrated by unique modular organization of polysaccharide-degrading enzymes. *Proc Natl Acad Sci* 111:E3708–E3717. <https://doi.org/10.1073/pnas.1406156111>.
26. Kawaguchi K, Senoura T, Ito S, Taira T, Ito H, Wasaki J, Ito S. 2014. The mannobiose-forming exo-mannanase involved in a new mannan catabolic pathway in *Bacteroides fragilis*. *Arch Microbiol* 196:17–23. <https://doi.org/10.1007/s00203-013-0938-y>.
27. Reddy SK, Bågenholm V, Pudlo NA, Bouraoui H, Koropatkin NM, Martens EC, Ståhlbrand H. 2016. A β -mannan utilization locus in *Bacteroides ovatus* involves a GH36 α -galactosidase active on galactomannans. *FEBS Lett* 590:2106–2118. <https://doi.org/10.1002/1873-3468.12250>.
28. La Rosa SL, Kachrimaniou V, Buffetto F, Pope PB, Pudlo NA, Martens EC, Rastall RA, Gibson GR, Westereng B. 2019. Wood-derived dietary fibers promote beneficial human gut microbiota. *mSphere* 4:e00554-18. <https://doi.org/10.1128/mSphere.00554-18>.
29. Tang K, Lin Y, Han Y, Jiao N. 2017. Characterization of potential polysaccharide utilization systems in the marine *Bacteroidetes* *Gramella flava* JLT2011 using a multi-omics approach. *Front Microbiol* 8:220. <https://doi.org/10.3389/fmicb.2017.00220>.
30. Larsbrink J, Rogers TE, Hemsworth GR, McKee LS, Tauzin AS, Spadiut O, Klinger S, Pudlo NA, Urs K, Koropatkin NM, Creagh AL, Haynes CA, Kelly AG, Cederholm SN, Davies GJ, Martens EC, Brumer H. 2014. A discrete genetic locus confers xyloglucan metabolism in select human gut *Bacteroidetes*. *Nature* 506:498–502. <https://doi.org/10.1038/nature12907>.
31. Broeksema B, Calusinska M, McGee F, Winter K, Bongiovanni F, Goux X, Wilmes P, Delfosse P, Ghoniem M. 2017. ICoVeR – an interactive visualization tool for verification and refinement of metagenomic bins. *BMC Bioinformatics* 18:233. <https://doi.org/10.1186/s12859-017-1653-5>.
32. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. DbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* 40:W445–W451. <https://doi.org/10.1093/nar/gks479>.
33. Busk PK, Pilgaard B, Lezyk MJ, Meyer AS, Lange L. 2017. Homology to peptide pattern for annotation of carbohydrate-active enzymes and prediction of function. *BMC Bioinformatics* 18:214. <https://doi.org/10.1186/s12859-017-1625-9>.
34. Galperin MY, Makarova KS, Wolf YI, Koonin EV. 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res* 43:D261–D269. <https://doi.org/10.1093/nar/gku1223>.
35. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* 428:726–731. <https://doi.org/10.1016/j.jmb.2015.11.006>.
36. Stewart RD, Auffret MD, Warr A, Wiser AH, Press MO, Langford KW, Liachko I, Snelling TJ, Dewhurst RJ, Walker AW, Roehe R, Watson M. 2018. Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. *Nat Commun* 9:870. <https://doi.org/10.1038/s41467-018-03317-6>.
37. Parks DH, Rinke C, Chuvochina M, Chaumeil P, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW. 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* 2:1533–1542. <https://doi.org/10.1038/s41564-017-0012-7>.
38. Saburi W. 2016. Functions, structures, and applications of cellobiose 2-epimerase and glycoside hydrolase family 130 mannoside phosphorylases. *Biosci Biotechnol Biochem* 80:1294–1305. <https://doi.org/10.1080/09168451.2016.1166934>.
39. Wang K, Yin J, Shen D, Li N. 2014. Anaerobic digestion of food waste for volatile fatty acids (VFAs) production with different types of inoculum: effect of pH. *Bioresour Technol* 161:395–401. <https://doi.org/10.1016/j.biortech.2014.03.088>.
40. Joyce A, Ijaz UZ, Nzeteu C, Vaughan A, Shirran SL, Botting CH, Quince C, O'Flaherty V, Abram F. 2018. Linking microbial community structure and function during the acidified anaerobic digestion of grass. *Front Microbiol* 9:540. <https://doi.org/10.3389/fmicb.2018.00540>.
41. Heintz-Buschart A, Wilmes P. 2018. Human gut microbiome: function matters. *Trends Microbiol* 26:563–574. <https://doi.org/10.1016/j.tim.2017.11.002>.
42. Zykwinska A, Rondeau-Mouro C, Garnier C, Thibault J-F, Ralet M-C. 2006. Alkaline extractability of pectic arabinan and galactan and their mobility in sugar beet and potato cell walls. *Carbohydr Polym* 65:510–520. <https://doi.org/10.1016/j.carbpol.2006.02.012>.
43. Watzlawick H, Morabbi Heravi K, Altenbuchner J. 2016. Role of the ganspqab operon in degradation of galactan by *Bacillus subtilis*. *J Bacteriol* 198:2887–2896. <https://doi.org/10.1128/JB.00468-16>.
44. Wefers D, Dong J, Abdel-Hamid AM, Müller Paul H, Pereira GV, Han Y, Dodd D, Baskaran R, Mayer B, Mackie RI, Canna I. 2017. Enzymatic mechanism for arabinan degradation and transport in the thermophilic bacterium *Caldanaerobius polysaccharolyticus*. *Appl Environ Microbiol* 83:e00794-17. <https://doi.org/10.1128/AEM.00794-17>.
45. Abbott DW, Hrynuk S, Boraston AB. 2007. Identification and characterization of a novel periplasmic polygalacturonic acid binding protein from *Yersinia enterocolitica*. *J Mol Biol* 367:1023–1033. <https://doi.org/10.1016/j.jmb.2007.01.030>.
46. Fujimoto Z, Jackson A, Michikawa M, Maehara T, Momma M, Henrissat B, Gilbert HJ, Kaneko S. 2013. The structure of a *Streptomyces avermitilis* α -L-rhamnosidase reveals a novel carbohydrate-binding module CBM67 within the six-domain arrangement. *J Biol Chem* 288:12376–12385. <https://doi.org/10.1074/jbc.M113.460097>.
47. Maus I, Koeck DE, Cibis KG, Hahnke S, Kim YS, Langer T, Kreubel J, Erhard M, Bremges A, Off S, Stolze Y, Jaenicke S, Goesmann A, Sczyrba A, Scherer P, König H, Schwarz WH, Zverlov VV, Liebl W, Pühler A, Schlüter A, Klocke M. 2016. Unraveling the microbiome of a thermophilic biogas plant by metagenome and metatranscriptome analysis complemented by characterization of bacterial and archaeal isolates. *Biotechnol Biofuels* 9:171. <https://doi.org/10.1186/s13068-016-0581-3>.
48. Alessi AM, Bird SM, Oates NC, Li Y, Dowle AA, Novotny EH, Deazevedo ER, Bennett JP, Polikarpov I, Young JPW, McQueen-Mason SJ, Bruce NC. 2018. Defining functional diversity for lignocellulose degradation in a microbial community using multi-omics studies. *Biotechnol Biofuels* 11:166. <https://doi.org/10.1186/s13068-018-1164-2>.
49. Esenther GR, Kirk TK. 1974. Catabolism of aspen sapwood in *Reticulitermes flavipes*. *Ann Entomol Soc Am* 67:989–991. <https://doi.org/10.1093/aesa/67.6.989>.
50. Li YF, Nelson MC, Chen PH, Graf J, Li Y, Yu Z. 2015. Comparison of the microbial communities in solid-state anaerobic digestion (SS-AD) reactors operated at mesophilic and thermophilic temperatures. *Appl Microbiol Biotechnol* 99:969–980. <https://doi.org/10.1007/s00253-014-6036-5>.
51. Guo J, Peng Y, Ni BJ, Han X, Fan L, Yuan Z. 2015. Dissecting microbial community structure and methane-producing pathways of a full-scale anaerobic reactor digesting activated sludge from wastewater treatment by metagenomic sequencing. *Microb Cell Fact* 14:33. <https://doi.org/10.1186/s12934-015-0218-4>.
52. Wieczorek AS, Hetz SA, Kolb S. 2014. Microbial responses to chitin and chitosan in oxic and anoxic agricultural soil slurries. *Biogeosciences* 11:3339–3352. <https://doi.org/10.5194/bg-11-3339-2014>.

53. Köhler T, Stingl U, Meuser K, Brune A. 2008. Novel lineages of Planctomycetes densely colonize the alkaline gut of soil-feeding termites (*Cubitermes* spp.). *Environ Microbiol* 10:1260–1270. <https://doi.org/10.1111/j.1462-2920.2007.01540.x>.
54. Limam RD, Chouari R, Mazéas L, Wu T-D, Li T, Grossin-Debattista J, Guerquin-Kern J-L, Saidi M, Landoulsi A, Sghir A, Bouchez T. 2014. Members of the uncultured bacterial candidate division WWE1 are implicated in anaerobic digestion of cellulose. *Microbiol Open* 3:157–167. <https://doi.org/10.1002/mbo3.144>.
55. Zhang C, Chen J, Yang F. 2014. Konjac glucomannan, a promising polysaccharide for OCDDS. *Carbohydr Polym* 104:175–181. <https://doi.org/10.1016/j.carbpol.2013.12.081>.
56. Mikkelsen A, Maaheimo H, Hakala TK. 2013. Hydrolysis of konjac glucomannan by *Trichoderma reesei* mannanase and endoglucanases Cel7B and Cel5A for the production of glucomannooligosaccharides. *Carbohydr Res* 372:60–68. <https://doi.org/10.1016/j.carres.2013.02.012>.
57. Chen J, Robb CS, Unfried F, Kappelman L, Markert S, Song T, Harder J, Avci B, Becher D, Xie P, Amann RL, Hehemann J-H, Schweder T, Teeling H. 2018. Alpha- and beta-mannan utilization by marine *Bacteroidetes*. *Environ Microbiol* 20:4127–4140. <https://doi.org/10.1111/1462-2920.14414>.
58. Naas AE, Mackenzie AK, Mravec J, Schückel J, Willats WGT, Eijssink VGH, Pope PB. 2014. Do rumen *Bacteroidetes* utilize an alternative mechanism for cellulose degradation? *mBio* 5:e01401-14. <https://doi.org/10.1128/mBio.01401-14>.
59. Temple MJ, Cuskin F, Baslé A, Hickey N, Speciale G, Williams SJ, Gilbert HJ, Lowe EC. 2017. A *Bacteroidetes* locus dedicated to fungal 1, 6–glucan degradation: Unique substrate conformation drives specificity of the key endo-1, 6–glucanase. *J Biol Chem* 292:10639–10650. <https://doi.org/10.1074/jbc.M117.787606>.
60. Zückert WR. 2014. Secretion of bacterial lipoproteins: through the cytoplasmic membrane, the periplasm and beyond. *Biochim Biophys Acta* 1843:1509–1516. <https://doi.org/10.1016/j.bbamcr.2014.04.022>.
61. Shipman JA, Berleman JE, Salyers AA. 2000. Characterization of four outer membrane proteins involved in binding starch to the cell surface of *Bacteroides thetaiotaomicron*. *J Bacteriol* 182:5365–5372. <https://doi.org/10.1128/JB.182.19.5365-5372.2000>.
62. Pugsley AP. 1993. The complete general secretory pathway in gram-negative bacteria. *Microbiol Rev* 57:50–108.
63. Tauzin AS, Laville E, Xiao Y, Nouaille S, Le Bourgeois P, Heux S, Portais JC, Monsan P, Martens EC, Potocki-Veronese G, Bordes F. 2016. Functional characterization of a gene locus from an uncultured gut *Bacteroides* conferring xylo-oligosaccharides utilization to *Escherichia coli*. *Mol Microbiol* 102:579–592. <https://doi.org/10.1111/mmi.13480>.
64. Pradeep GC, Cho SS, Choi YH, Choi YS, Jee JP, Seong CN, Yoo JC. 2016. An extremely alkaline mannanase from *Streptomyces* sp. CS428 hydrolyzes galactomannan producing series of mannoooligosaccharides. *World J Microbiol Biotechnol* 32:84. <https://doi.org/10.1007/s11274-016-2040-5>.
65. Senoura T, Ito S, Taguchi H, Higa M, Hamada S, Matsui H, Ozawa T, Jin S, Watanabe J, Wasaki J, Ito S. 2011. New microbial mannan catabolic pathway that involves a novel mannosylglucose phosphorylase. *Biochem Biophys Res Commun* 408:701–706. <https://doi.org/10.1016/j.bbrc.2011.04.095>.
66. Kawahara R, Saburi W, Odaka R, Taguchi H, Ito S, Mori H, Matsui H. 2012. Metabolic mechanism of mannan in a ruminal bacterium, *Ruminococcus albus*, involving two mannoside phosphorylases and cellobiose 2-epimerase: discovery of a new carbohydrate phosphorylase, β -1,4-mannoooligosaccharide phosphorylase. *J Biol Chem* 287:42389–42399. <https://doi.org/10.1074/jbc.M112.390336>.
67. Bagenholm V, Reddy SK, Bouraoui H, Morrill J, Kulcinskaja E, Bahr CM, Aurelius O, Rogers T, Xiao Y, Logan DT, Martens EC, Koropatkin NM, Stalbrand H. 2017. Galactomannan catabolism conferred by a polysaccharide utilization locus of *Bacteroides ovatus*: Enzyme synergy and crystal structure of a β -mannanase. *J Biol Chem* 292:229–243. <https://doi.org/10.1074/jbc.M116.746438>.
68. Pawar P-A, Koutaniemi S, Tenkanen M, Mellerowicz EJ. 2013. Acetylation of woody lignocellulose: significance and regulation. *Front Plant Sci* 4:118. <https://doi.org/10.3389/fpls.2013.00118>.
69. Segata N, Börnigen D, Morgan XC, Huttenhower C. 2013. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat Commun* 4:2304. <https://doi.org/10.1038/ncomms3304>.
70. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <https://doi.org/10.1186/1471-2105-11-119>.
71. Terrapon N, Lombard V, Drula É, Lapébie P, Al-Masaudi S, Gilbert HJ, Henrissat B. 2017. PULDB: the expanded database of polysaccharide utilization loci. *Nucleic Acids Res* 46:D677–D683. <https://doi.org/10.1093/nar/gkx1022>.
72. Busk PK, Lange M, Pilgaard B, Lange L. 2014. Several genes encoding enzymes with the same activity are necessary for aerobic fungal degradation of cellulose in nature. *PLoS One* 9:e114138. <https://doi.org/10.1371/journal.pone.0114138>.
73. Markowitz VM, Chen IMA, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, Huntemann M, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC. 2012. IMG: The integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res* 40:D115–D122. <https://doi.org/10.1093/nar/gkr1044>.
74. Jaffee MB, Imperiali B. 2013. Optimized protocol for expression and purification of membrane-bound PglB, a bacterial oligosaccharyl transferase. *Protein Expr Purif* 89:241–250. <https://doi.org/10.1016/j.pep.2013.04.001>.
75. Tiralongo J, Maggioni A. 2010. The targeted expression of nucleotide sugar transporters to the *E. coli* inner membrane. *Methods Mol Biol* 705:237–249. https://doi.org/10.1007/978-1-61737-967-3_14.
76. Juncker AS, Willenbrock H, Heijne G, Brunak S, Nielsen H, Krogh A. 2003. Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci* 12:1652–1662. <https://doi.org/10.1110/ps.0303703>.
77. Rouland C, Civas A, Renoux J, Petek F. 1988. Synergistic activities of the enzymes involved in cellulose degradation purified from *Macrotermes mülleri* and from its symbiotic fungus *Termitomyces* sp. *Comp Biochem Physiol* 91:459–465.
78. Rouland C, Civas A, Renoux J, Petek F. 1988. Purification and properties of cellulases from the termite *Macrotermes mülleri* and its symbiotic fungus. *Comp Biochem Physiol* 91:449–458. [https://doi.org/10.1016/0305-0491\(88\)90004-1](https://doi.org/10.1016/0305-0491(88)90004-1).
79. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. <https://doi.org/10.1186/1471-2164-9-75>.