

# Gut microbiome of capybara, the Amazon master of the grasses, harbors unprecedented enzymatic strategies for plant glycans breakdown

**Lucelia Cabral**

CNPEM: Centro Nacional de Pesquisa em Energia e Materiais

**Gabriela F. Persinoti**

CNPEM: Centro Nacional de Pesquisa em Energia e Materiais

**Douglas A. A. Paixao**

CNPEM: Centro Nacional de Pesquisa em Energia e Materiais

**Marcele P. Martins**

CNPEM: Centro Nacional de Pesquisa em Energia e Materiais

**Mariana Chinaglia**

CNPEM: Centro Nacional de Pesquisa em Energia e Materiais

**Mariane N. Domingues**

CNPEM: Centro Nacional de Pesquisa em Energia e Materiais

**Mariana A. B. Morais**

CNPEM: Centro Nacional de Pesquisa em Energia e Materiais

**Mauricio L. Sforca**

CNPEM: Centro Nacional de Pesquisa em Energia e Materiais

**Renan A. S. Pirolla**

CNPEM: Centro Nacional de Pesquisa em Energia e Materiais

**Wesley C. Generoso**

CNPEM: Centro Nacional de Pesquisa em Energia e Materiais

**Clelton A. Santos**

CNPEM: Centro Nacional de Pesquisa em Energia e Materiais

**Lucas F. Maciel**

CNPEM: Centro Nacional de Pesquisa em Energia e Materiais

**Nicolas Terrapon**

Aix-Marseille Université: Aix-Marseille Université

**Vincent Lombard**

Aix-Marseille Université: Aix-Marseille Université

**Bernard Henrissat**

Aix-Marseille Université: Aix-Marseille Université

**Mario T. Murakami** (✉ [mario.t.murakami@gmail.com](mailto:mario.t.murakami@gmail.com))

## Research Article

**Keywords:** capybara, gut microbiome, multi-omics, CAZymes, plant polysaccharides.

**Posted Date:** April 29th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-456076/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background:** Plant biomass is a promising feedstock to replace fossil-based products including fuels, chemicals and materials. However, the high resistance of plant biomass to either physicochemical or biological deconstruction has been hampering its broad industrial utilization and, consequently, the transition to a sustainable bioeconomy. The gut system from herbivores are formidable bioreactors in nature for lignocellulose breakdown and the diverse ecological niches where herbivores are found have led to the rise of a myriad of molecular strategies to cope with the sheer complexity of plant polysaccharides. This study illuminates how the underexplored microbiota of the largest living rodent, capybara, found in Pantanal wetlands and the Amazon basin, can efficiently depolymerize and utilize lignocellulosic biomass.

**Results:** Here, we have elucidated the gut microbial structure and composition of the semiaquatic herbivorous capybara through multi-omics approaches. Metabolic reconstruction of this microbiota showed that cellulose degradation is likely performed by *Fibrobacter* bacteria, whereas hemicelluloses and pectins are processed by a broad arsenal of Carbohydrate-Active enZymes (CAZymes) organized in polysaccharide utilization loci (PULs) identified in the multiple metagenome-assembled genomes from the phylum *Bacteroidetes*. Furthermore, metabolomics analysis showed short chain fatty acids as major fermentation products, which are key markers of digestion performance of plant polysaccharides. Exploring the genomic dark matter of this gut microbial community, two novel CAZymes families were unveiled including a glycoside hydrolase family of  $\beta$ -galactosidases (GHXXX) and a carbohydrate-binding module family (CBMXX) involved in xylan binding that establishes an unprecedented three-dimensional fold among associated modules to CAZymes.

**Conclusions:** Our results reveal how the capybara gut microbiota orchestrates the depolymerization and utilization of dietary plant polysaccharides, representing an untapped reservoir of new and intricate enzymatic strategies to overcome the recalcitrance of plant polysaccharides, a central challenge toward a circular and sustainable economy.

## Background

The diverse symbiotic microbiota within the digestive tract of herbivores has been an overwhelming source of diverse enzymatic mechanisms for lignocellulose deconstruction [1–4]. For decades, the microbiota of foregut (rumen) fermenters has been employed as a model system [5, 6], which led to the discovery of sophisticated and fascinating systems to degrade complex plant fibers, such as the multi-enzyme complexes (cellulosomes) from *Ruminococcus flavefaciens* [7] and the efficient and distinguishing cellulose degradation system from *Fibrobacter succinogenes* [8].

A less explored and equally effective class of herbivores is the hindgut fermenters [9]. Similar to foregut fermenters, the digestion is accomplished by a symbiotic microbial community, but in a single and enlarged fermentation chamber [10]. These monogastric herbivores comprise a vast range of animals,

from massive mammals, such as elephants, rhinos, and horses, to small animals, as rabbits and semiaquatic rodents [11]. In addition, they are spread over a myriad of ecological niches, which suggests that highly specialized molecular strategies may have emerged to overcome the complexity and diversity of plant glycans in these environments.

The capybara (*Hydrochoerus hydrochaeris*) is the largest living rodent, typically found in the Pantanal wetlands and Amazon basin, and it is also known as “the master of the grasses” due to its diet based on gramineous and aquatic plants. In this animal, the fermentation takes place in the cecum that corresponds to almost three quarters of the gastrointestinal tract, reaching a digestive efficiency comparable to that of ruminants [12]. Moreover, as a strategy to maximize absorption of nutrients derived from bacterial fermentation, capybara can eat their cecotropes, a specific type of soft excreta [13]. This habit of cecotrophy is more frequent in wild animals during dry season, when food is scarce [13]. Despite being a formidable plant biomass fermenter, the enzymatic strategies and metabolic pathways employed by its microbial symbiotic community for the breakdown and utilization of recalcitrant dietary fibers remain mostly elusive. In addition, wild capybara animals dwelling the Southeast region of Brazil have incorporated sugarcane in their diet for decades [14], which makes their cecal microbiome an especially attractive system for lignocellulose depolymerization of this industrially relevant feedstock.

To elucidate the enzymatic strategies employed by the Brazilian capybara microbiota for plant cell wall deconstruction, we comprehensively investigated this gut microbial community combining an integrated multi-omics approach (16S rRNA gene targeting sequencing, metagenomics, metatranscriptomics and NMR-based metabolomics) with carbohydrate enzymology and X-ray crystallography, which ultimately led to the discovery of two novel CAZy families. These findings highlight the potential of the capybara gut microbiome as a reservoir of unprecedented enzymatic systems for carbohydrate processing, and thereby expanding our current understanding of gut microbial strategies to overcome the lignocellulose recalcitrance, which might be instrumental to foster the development of bio-based technologies.

## Results

### Taxonomic distribution of the capybara gut microbiota diverges from typical hindgut fermenters

The taxonomic structure of the capybara gut microbiota is mainly constituted by *Bacteria*, with only a small fraction of reads corresponding to *Archaea* (16S: 1.07%, MG: 0.40% and MT: 1.55%) and *Fungi* (MG 0.12% and MT: 0.32%). 16S rRNA gene-based taxonomy analysis indicates that the most abundant bacteria found in capybara gut microbiota are members from the phyla *Firmicutes* (mean  $\pm$  sd:  $35.8 \pm 12.4\%$ ), *Bacteroidetes* ( $31.5 \pm 9.8\%$ ), followed by *Fusobacteria* ( $15.3 \pm 5.4\%$ ) and *Proteobacteria* ( $8.4 \pm 5.4\%$ ) (Fig. 1a). A similar taxonomic distribution was observed for 16S rRNA reads recovered from metagenome (16S\_MG), MG and MT datasets (correlation coefficient  $r=0.96$  for MG:MT,  $r=0.74$  for 16S\_MG:MG and  $r=0.71$  for 16S\_MG:MT,  $P < 0.05$ ). *Bacteroidetes*, *Fusobacteria* and *Proteobacteria* phyla were significantly more abundant in MG than in MT considering cecal samples (Fig. 1b), whereas *Euryarchaeota* (MT/MG ratio expressed as mean  $\pm$  SD:  $2.3 \pm 0.6$ ), *Fibrobacteres* ( $1.7 \pm 0.9$ ) and

*Spirochaetes* (1.6 +/- 0.4) were more prevalent in MT than in MG data of cecal samples (Fig. 1b). Although the protein-to-RNA ratios are rather dynamic in complex microbial communities according to absolute multi-omics studies [15], the higher MT/MG ratios for some phyla, such as *Euryarchaeota* known to include methanogens species, might correlate to the fact that capybara is a pronounced methane producer among the hindgut fermenters [12, 16].

Binning of MG assembled contigs based on tetranucleotide frequency and coverage profile resulted in the reconstruction of 79 unique Metagenome-Assembled Genomes (MAGs) (Suppl. Table S1), being 24 considered of high quality (completeness >90% and contamination <5%) and 50 medium-quality (completeness >50% and contamination <10%), in agreement with the completeness and contamination parameters suggested by Bowers *et al.* 2017 [17]. Based on the GTDB database, 24 of the recovered MAGs are classified as taxonomic novelties, representing either novel species or genera, which include one from *Fibrobacterota* (family *Fibrobacteraceae*), one from *Planctomycetota* (family *Thermoguttaceae*), one from *Spirochaetota* (family *Sphaerochaetaceae*), eight from *Firmicutes* (families *Erysipelotrichaceae*, *Lachnospirales*, *CAG-826* and *Oscillospiraceae*) and 13 from *Bacteroidetes* (families *Bacteroidaceae*, *UBA932* and *Muribaculaceae*) (Suppl. Table S1). It is notable that despite only two MAGs from *Fusobacteria* and *Proteobacteria* (MAG38, 39 and MAG77,78, respectively) were recovered from MG, the most abundant OTUs revealed by 16S rRNA gene analysis were classified to these phyla (20% *Fusobacteria* and 18% *Proteobacteria*, Suppl. Fig. S1), pointing to an important role of these species in this microbiota.

A prevalence of *Firmicutes*, *Proteobacteria*, *Bacteroidetes* and *Tenericutes* was observed in the gut microbiomes of other hindgut herbivores such as *Castor fiber*, *Castor canadensis*, horse, rabbit, and koala [18–22]. Furthermore, microbiota analysis of domesticated herbivores including hindgut fermenters, ruminants and monogastric animals revealed *Firmicutes* as the most abundant phylum (53.11, 63.35 and 52.27%, respectively), followed by *Bacteroidetes* (31.36, 20.95 and 26.95%, respectively) [23]. Although the predominance of *Bacteroidetes* and *Firmicutes* is a general feature of mammalian gut microbiomes [24], the microbiota of native Brazilian capybara differs from other hindgut fermenters and ruminants, mainly due to a reduced abundance of *Firmicutes* (35%) along with a higher abundance of *Fusobacteria* (15%) and *Proteobacteria* (8%), suggesting that diversified strategies for lignocelulosic utilization may have evolved in this gut environment.

### ***Fibrobacteres* and *Bacteroidetes* are main degraders of dietary fibers in capybara gut**

In order to understand the ability of capybara to convert plant polysaccharides into free sugars, MG and MT data were investigated to determine the genomic potential of its gut microbiota associated with Carbohydrate-Active enZymes (CAZymes). A total of 6,132 putative CAZymes genes encoding for 105 Glycoside Hydrolases (GH), 11 Carbohydrate Esterases (CE), and 10 Polysaccharide Lyases (PL) families were identified, of which 456 genes presented a modular architecture (Suppl. Table S2). The most abundant CAZymes identified are members of the families GH3, GH2 and GH1 (by decreasing abundance) (Fig. 2), which is in agreement with that reported for other host-associated gut microbiomes

such as human, swine and cattle rumen [25]. These enzymes encompass diversified activities including  $\beta$ -glucosidase,  $\beta$ -xylosidase,  $\beta$ -galactosidase and  $\beta$ -mannosidase, and are often associated with the final steps in the depolymerization cascade of several plant polysaccharides such as cellulose, heteroxylans, mixed-linkage  $\beta$ -glucans and  $\beta$ -mannans.

In the CAZyme repertoire of this microbiota neither cellulases from families GH6, GH7 and GH48, nor cellulosomes, assessed by the presence of cohesin and dockerin domains associated with cellulases, could be identified in MG or MT datasets. In ruminal anaerobic fungi, these families are found in high abundance, possibly targeting recalcitrant cellulose structures [26]. However, in the capybara gut microbiota assayed here, fungi were detected only at very low abundance (Fig. 1a). This suggests that cellulose degradation in the capybara gut might be mainly accomplished by endo- $\beta$ -1,4-glucanases (EC 3.2.1.4) from families GH5 (subfamilies GH5\_2, GH5\_4, GH5\_25 and GH5\_37), GH8, GH9 and GH45, which were detected either as single domains or in multi-modular protein architectures. Notably, the most expressed genes encoding endo- $\beta$ -1,4-glucanases belong to families GH5\_2, GH8, GH9 and GH45, and were identified in *Fibrobacter* MAGs (Suppl. Fig. S2 and Suppl. Table S3) that also present a high MT/MG ratio (Fig. 1b), indicating that these bacteria are putatively key players in cellulose degradation in the capybara gut. *Fibrobacter succinogenes* is known as a highly efficient cellulolytic bacterium in the cow rumen [27] and employs a multi-protein complex to attach to cellulose fibers and cellulases secreted by the T9SS-dependent secretion system for cellulose breakdown [28]. The three *Fibrobacter* MAGs recovered from capybara gut microbiome encode cellulases with a T9SS signal sequence as well as proteins for cellulose adhesion including tetratricopeptide, fibro-slime, OmpA and pilin proteins, as reported for *F. succinogenes* [28]. Furthermore, from the set of 347 proteins observed in the outer membrane vesicles (OMVs) from *F. succinogenes* [29], we have identified 262 with sequence identity ranging from 30-99%. These observations suggest that typical *Fibrobacter* mechanisms, fundamentally relying on cell surface adhesion and OMVs, are central for cellulose depolymerization in the capybara gut.

In the multiple recovered *Bacteroidetes* MAGs, a large number of polysaccharide utilization loci (PULs) and clusters of CAZymes (CCs) were identified (Fig. 3 and Suppl. Table S4), which provide highly diversified capabilities to this microbiota to cope with the chemical and structural complexity of typically abundant hemicelluloses and pectins in gramineous or aquatic plants such as heteroxylans, mixed-linkage  $\beta$ -glucans,  $\beta$ -1,3-glucans, xyloglucans, mannans and homogalacturonans. Notably, the identified PUL targeting mixed-linkage  $\beta$ -glucans (Suppl. Fig. S3) is highly conserved in capybara and human microbiomes, presenting identical gene architecture encompassing one GH16 and two GH3 enzymes, as reported in [30]. PULs targeting to heteroxylans and homogalacturonans (Suppl. Fig. S3), common components of gramineous plants such as sugarcane [31], also resemble PULs identified in human gut microbiomes [32], highlighting the significant level of conservation of *Bacteroidetes* enzymatic systems in omnivores and hindgut herbivores.

Despite the presence of multiple carbohydrate esterases (CEs) (Fig. 3), the lack of auxiliary-active enzymes (AAs) indicates a low capacity of the capybara microbiota to perform plant biomass delignification as also observed for other monogastric herbivores such as horses [33, 34]. As a

mechanism to cope with lignin-rich diets, these animals may employ cecotrophy to enhance digestibility and nutrient uptake. In addition, it is noteworthy that many identified PULs only showed similarity with non-experimentally validated PULs and without a clear substrate target (Suppl. Table S4), which in part could be due to intrinsic limitations of genome reconstruction from metagenomes, but also reflects the variability, heterogeneity and partial knowledge of the structure and composition of the glycans present in the diet of wild capybaras.

Taken together, the CAZyme analysis of the capybara gut microbiota indicates *Fibrobacteres* as main drivers for cellulose breakdown, whereas the numerous *Bacteroidetes* PULs/CCs confer to this community a myriad of enzymatic strategies to tackle with the complex and diverse hemicelluloses and pectins typically present in gramineous and aquatic plants, major components of capybara diet.

### **Global metabolite profiling shows high performance on the conversion of dietary fibers into short-chain fatty acids**

Once addressed important players in the depolymerization of dietary fibers in capybara gut, we further investigated the role of these microorganisms in the conversion of free sugars into energy for the host by integrating metabolomics and metabolic reconstruction analysis.

The major fermentation products measured in the capybara gut were short-chain fatty acids (SCFAs), among more than 40 metabolites detected by NMR spectroscopy-based metabolomics (Suppl. Table S5). The most abundant metabolites observed in cecal and rectal samples were acetate (mean  $\pm$  SD:  $74.83 \pm 22.17$  and  $30.40 \pm 22.76$  mM, respectively), propionate ( $31.0 \pm 6.67$  and  $15.98 \pm 12.8$  mM) and butyrate ( $23.30 \pm 5.63$  and  $8.35 \pm 12.83$  mM). These SCFA ratios indicate a forage-based diet and are similar to that seen for ruminants [35, 36], supporting a high efficiency of this microbiota in the use of dietary fibers as an energy source.

Genes related to pyruvate fermentation into acetate were highly abundant in both MG and MT data for cecal and rectal samples, and they are derived from *Firmicutes*, *Bacteroidetes* and *Fusobacteria* (Fig. 4 and Suppl. Fig. S4). Metabolic pathway reconstruction analysis shows that acetate can be putatively produced by any of the bacterial MAGs recovered from capybara gut microbiome (Fig. 4), which is in agreement with the high abundance of this metabolite in both cecal and rectal samples (Suppl. Table S5). On the other hand, the expression analysis of key genes involved in the butyrate pathway (*atoA/D* genes) indicates that *Firmicutes* *Ileibacterium* sp. MAG6 and *Megasphaera* sp. MAG33 are likely the major butyrate-producing bacteria in the capybara gut (Suppl. Fig. S5 and Suppl. Table S6). The *Bacteroidetes* *Marinilabiliaceae* MAG47 and *Fusobacteria* MAG38 and MAG39 also have co-localized genes *atoA/atoD* and *ptb/butK*, suggesting that they also contribute to butyrate production, in some extent (Suppl. Fig. S5 and Suppl. Table S6). The typical genes from acrylate and propanediol pathways involved in propionate production were not identified in the recovered MAGs from capybara gut (Fig. 4), but the *mmdA* gene encoding a methylmalonyl-CoA decarboxylase from the succinate pathway, is widespread mainly among *Bacteroidetes* and was also observed in some *Firmicutes* and *Fusobacteria* MAGs (Suppl. Fig. S5 and Suppl. Table S6). Furthermore, the ratio of propionate detected in the gut

capybara gut correlates ( $R=0.77$  and  $p=0.07$ ) with the relative abundance of *Bacteroidetes*, supporting the succinate pathway from this phylum as the major source of propionate production in the capybara gut.

Together, these results demonstrate a high SCFA production in the capybara digestive tract, which is a common marker of digestion performance of dietary fibers [37], and therefore, reinforce the potential of this microbiota for the breakdown of recalcitrant plant polysaccharides with concomitant production of energetic metabolites for the host.

### **A new GH family mined from the genomic dark matter of capybara microbiome**

Drawing on the results showed herein, capybara gut microbiome can be an important source for uncharted enzymes involved in plant polysaccharides depolymerization. Moreover, the joint MG and MT analysis of capybara gut microbiome revealed several expressed genes annotated as hypothetical proteins. Some of these genes are remotely similar to CAZy members, with sequence identity ranging from 10 to 20%, suggesting a potential function in the processing of plant polysaccharides, but requiring further functional investigation (Suppl. Table S7). Aiming to uncover the activity of these proteins, several ORFs were expressed and subjected to biochemical assays employing a diverse set of synthetic, poly- and oligosaccharides substrates (Suppl. Table S8).

One of these proteins (SEQ ID PBMDCECB\_44807, named here CapGHXXX) was active on *p*-nitrophenyl- $\beta$ -D-galactopyranoside (*p*NP- $\beta$ -D-Gal) and kinetic parameters were determined from substrate saturation curves (Table 1 and Suppl. Fig. S6). CapGHXXX orthologues are found in *Actinobacteria*, *Firmicutes*, *Verrucomicrobia* and *Bacteroidetes* MAGs recovered from diverse sources such as rumen, feces, gut, and oral microbiotas (Suppl. Table S9), being the closest sequence from a rumen-derived MAG (UBA2817) from the uncultured *RC9* group [38]. Sequence analysis showed that CapGHXXX is distantly related to families GH5 and GH30 (Fig. 5a) and protein threading indicates a TIM-barrel fold (Suppl. Fig. S7), suggesting that this GH family belongs to the clan GH-A. To further explore the GHXXX family, the enzyme CBK67650.1 (SEQ\_ID BXY\_26070) from *B. xylanisolvens*, which shares 46% sequence identity with CapGHXXX, was heterologously produced and biochemically characterized (Table 1). This second member also showed  $\beta$ -galactosidase activity that strengthens at biochemical level the establishment of this new GH family.

In the *Bacteroidota bacterium* MAG42 recovered from Capybara gut, CapGHXXX is found in a predicted PUL, additionally comprising enzymes from families GH2 and GH78. A similar PUL organization was also predicted in *Bacteroidetes* sp. 1\_1\_30 recovered from human gut, which yet harbors enzymes from GH36, CE7 and PL8\_2 families. It is noteworthy that CapGHXXX is often found fused to a GH36 module or in PULs having GH36 members, as in *B. xylanisolvens* and *Prevotella dentalis*, recovered from stool and oral cavity, respectively (Fig. 5b), indicating a synergistic relationship between these families. Moreover, these families are also commonly found along with GH78  $\alpha$ -L-rhamnosidases in the PUL context. In *Bacteroidales bacterium* UBA2817, a GHXXX member is appended to a GH78 module carrying a CBM67, both targeting rhamnogalacturonans (Fig. 5b). These observations suggest that GHXXX could act on  $\beta$ -linked galactosyl residues in pectic polysaccharides.

## Capybara *Prevotella* sp. MAG57 harbors a novel family of carbohydrate-binding module

Among the recovered MAGs from capybara gut microbiome, *Prevotella* sp. MAG57 is the one with the largest number of CAZyme-encoding genes. Phylogenetic and whole genome analyses show that MAG57 is closely related to other uncultured MAGs taxonomically assigned to the *Prevotellaceae* family recovered from the UBA project [38], either from sheep, elephant and mice gut microbiomes (Suppl. Fig. S8).

Multiple PULs and CAZyme clusters were identified in MAG57 (Suppl. Table S4) including a gene cluster targeting arabinoxylan (CC102), an abundant hemicellulose in secondary cell walls of sugarcane and other grasses. This cluster encodes two exo-enzymes from families GH43 and GH97, and an unconventional GH10 member with an unknown 45 kDa N-terminal domain (Fig. 6a). Sequence analysis showed that this unusual N-terminal domain is also present in *Bacteroidetes* MAGs derived from the gut of human, mouse, and elephant (Suppl. Table S10); however, it displays no similarity with any known ancillary domain associated with CAZymes.

Therefore, to evaluate the function of this unconventional GH10 member (CapGH10), the full-length protein, its domains apart, and the other GH members comprising the CC102 cluster were recombinantly expressed and characterized. The GH97 member (CapGH97) is a calcium-activated  $\alpha$ -galactosidase, whereas the GH43 member (CapGH43\_12) is a highly active  $\alpha$ -L-arabinofuranosidase (Suppl. Figs. S9-S10 and Table 1) – two key activities to remove decorations of heteroxylans.

The GH10 domain of CapGH10 exhibits endo- $\beta$ -1,4-xylanase activity, being active on both xylan and distinct arabinoxylans (Table 1). Kinetic analysis indicate that decorations present in rye arabinoxylan (arabinose/xylose ratio = 40/60) are not detrimental to the catalytic performance, showing similar  $K_m$  and  $k_{cat}$  constants compared to xylan (Table 1 and Suppl. Fig. S11). The Xyn10Z enzyme from *Hungateiclostridium thermocellum* ATCC 27405, sharing 36% of sequence identity with CapGH10, is the closest characterized member so far, with high activity on xylan [39]. The N-terminal region of Xyn10Z encompasses a feruloyl esterase followed by a CBM6 domain, which is not conserved in CapGH10 [40]. The CapGH10 N-terminus showed only sequence similarity with uncharacterized proteins, with the closest homologs mostly presenting a GH10 domain with sequence identity around 37-44%. CapGH10 orthologues, featuring similar domain architecture, were identified in PULs from ruminal *Prevotella* sp. such as *Prevotella* sp. BP1-148, *Prevotella* sp. BP1-145, *Prevotellaceae bacterium* HUN156 and *Prevotellaceae bacterium* MN60, also likely targeting xylan-related polysaccharides.

The potential enzymatic activity of the isolated N-terminal domain of CapGH10 was assessed against 30 different substrates including synthetic substrates, oligosaccharides, and polysaccharides (Suppl. Table S8), but no (hydrolase, lyase or esterase) activity was observed. Typical activities involved in heteroxylans breakdown including endo- $\beta$ -1,4-xylanase,  $\beta$ -xylosidase,  $\alpha$ -L-arabinofuranosidase,  $\alpha$ -D-galactosidase,  $\alpha$ -D-glucuronidase, 4-O-methyl-glucuronoyl methylesterase, feruloyl esterase and acetyl xylan esterase were assayed by distinct methods without the detection of product formation or substrate consumption. Under

this perspective, we then interrogated the capacity of this N-terminal domain to bind potential substrates of its GH10 partner such as beechwood xylan and arabinoxylans using affinity gel electrophoresis (AGE). As shown in Fig. 6c, this domain can indeed interact with the substrates of the GH10 domain, suggesting that this N-terminal domain may target the CapGH10 catalytic domain to xylan polysaccharides.

To get further insights into the potential role of this unconventional N-terminal domain, its crystallographic structure was solved by SeMet phasing at 1.8 Å resolution (Suppl. Table S11). The domain is monomeric in solution (Suppl. Fig. S12) and exhibits a parallel right-handed  $\beta$ -helix fold, consisting of 14 complete helical turns with two main short helices protruding from the  $\beta$ -helix backbone (Fig. 6b). The 14 helical turns are twisted and curved with a calcium ion between the 11<sup>th</sup> and 12<sup>th</sup> turns in an octahedral coordination sphere (Fig. 6b). This  $\beta$ -helix fold is observed in the clan GH-N of the GH superfamily, in the carbohydrate esterase CE8 and in several polysaccharide lyase (PL) families; however, structural comparisons with these CAZy families (GH28, GH91, PL6 and CE8) led to high rmsd values (> 3 Å), indicating poor three-dimensional conservation (Suppl. Table S12). Neither the catalytically relevant residues nor the active site topology of these families are conserved in the CapGH10  $\beta$ -helix domain (Suppl. Fig. S13), supporting that this domain is not catalytically active.

In order to elucidate the molecular determinants for xylan binding observed in AGE experiments, two surface regions populated with aromatic and acidic residues, typical platforms for carbohydrate interaction, were identified and mutated. The region I between the turns 1-4 and the region II near to turns 6-10 (Fig. 6b and Suppl. Fig. S14). Mutations E247A and E282A, in the region II, severely impaired protein stability and led to the expression only as inclusion bodies. Mutation D344L (at the calcium-binding site) also affected protein stability in a less extent, but the arabinoxylan/xylan binding capacity was preserved (Suppl. Fig. S15). This result indicates that calcium ion has a structural relevance rather than a functional role in carbohydrate recognition. Only the mutants Y62A and E82A, affected the migration pattern in AGE assays with beechwood xylan and rye arabinoxylan (Fig. 6c). Both residues are located at the region I, indicating that this patch plays a role in carbohydrate binding. It is worth to mention that two aromatic residues located at the corresponding region of the GH28 active site, Y193 and Y279, did not alter the carbohydrate binding, in agreement with no functional relevance of this region for CapGH10  $\beta$ -helix domain. Combining the biochemical, structural and mutagenesis analyses, we would define CapGH10  $\beta$ -helix domain as a CBM, therefore, establishing a novel structural scaffold in this superfamily and founding the family CBMXX.

Taken together, we conclude that this unprecedented modular endo- $\beta$ -1,4-xylanase along with the synergistic activities of other CC107 partners confers the ability to *Prevotella* sp. MAG57 to act on complex heteroxylans (Fig. 6d), a key function in the gut microbiome of capybara that have grasses as a major component in its diet.

## Discussion

Gut microbiota of herbivores is one of the most efficient systems in nature to overcome the recalcitrance of plant cell walls and the understanding of enzymatic and metabolic mechanisms employed by these microbial communities can provide novel and alternative solutions for the valorization of lignocellulosic materials and open new opportunities for carbohydrate-based biotechnological applications. In this study, we reveal how the gut microbiota of the largest living rodent, capybara (*Hydrochoerus hydrochaeris*) and also known as “master of the grasses”, can efficiently depolymerize and utilize plant polysaccharides. These semi-aquatic animals are hindgut fermenters throughout found in Pantanal wetlands and the Amazon basin, which have incorporated sugarcane in their diet for decades, raising the possibility that their gut microbiota has been shaped by this biomass of great industrial relevance.

The metagenomics, metatranscriptomics and CAZyme inventory analyses revealed that cellulose degradation in this community is not accomplished by classical mechanisms involving cellobiohydrolases or cellosomes. Instead, cellulose is likely processed by singular and sophisticated mechanisms featured by *Fibrobacteres*, including single and multi-modular CAZymes secreted by the T9SS system, CAZyme-rich outer membrane vesicles and lignocellulose adhesion proteins (Fig. 7). The complex and diverse architecture and composition of hemicellulosic and pectic polysaccharides present in gramineous and aquatic plants are tackled by a vast number of CAZymes organized in PULs found in the multiple recovered *Bacteroidetes* MAGs, which in part resembles to that from human gut *Bacteroidetes* species such as the PULs for mixed-linkage b-glucans [30] and xyloglucans [41].

It was prominent the identification of genes or PULs with remote or no similarity to known CAZy families and systems that led to the discovery of two new CAZymes families including a high-molecular weight CBM family involved in xylan recognition (CBMXX) and a GH-A clan family of b-galactosidases (GHXXX). This novel CBM family exhibits an unprecedented structural scaffold in this superfamily and its molecular architecture consisting of repeating b-helix units could serve as a platform for the rational design and engineering of CBMs. The GHXXX family expands the panel of industrially relevant CAZy families since b-galactosidases are broadly employed in food and beverage industries especially in the processing of dairy-based products.

The CAZyme repertoire of the capybara gut microbiome utterly covers the most abundant and recalcitrant polysaccharides present in gramineous and aquatic plants, which also requires efficient metabolic capabilities to further convert these depolymerized polysaccharides into SCFAs, the main energy source of the host. This hypothesis was further validated by global metabolite profiling and metabolic reconstructions with the three SCFAs (acetate, butyrate and propionate) as the main metabolites produced mostly by *Bacteroidetes*, *Firmicutes* and *Fusobacteria* (Fig. 7). Similar microbial strategies for SFCAs production were also observed in human gut bacteria, highlighting the commonalities between the gut microbiota from omnivores and hindgut fermenters [42]. Taken together, these results shed light on the molecular mechanisms of carbohydrate processing and metabolism by the native capybara gut microbiota.

## Conclusions

This work provides an unprecedented and comprehensive understanding of the enzymatic apparatus and metabolic pathways employed by the gut microbiota from the Amazon monogastric semi-aquatic herbivore, capybara, for the breakdown and utilization of recalcitrant dietary polysaccharides. This microbial community combines the unique cellulolytic machinery featured by *Fibrobacteres* and the diverse and elaborated PULs found in *Bacteroidetes* to efficiently depolymerize lignocellulosic biomass. Structural and functional investigation of proteins and PULs identified in the genomic dark matter of this microbiota uncovered two new CAZy families, highlighting its great potential as source of enzymes for the processing of plant polysaccharides. These findings expand our current understanding about gut microbial strategies to overcome the recalcitrance of lignocellulosic biomass, which might be utilized in biorefineries for the valorization of agroindustrial residues.

## Methods

### Ethics statement

This study was carried out in strict accordance with the Animal Management Rule of the Brazilian Ministry of Environment (Sisbio 59826-1). The samples were obtained from three euthanized animals in Tatuí/São Paulo State, Brazil (September 2017) as a measure of management of Rocky Mountain Spotted Fever (RMSF) hosts. After euthanasia, 20 g of intestinal contents were collected from the cecum and recto of each animal. All samples were placed in sterile containers and immediately frozen in liquid nitrogen. Samples were kept at  $-80^{\circ}\text{C}$  until processing.

### Microbial DNA extraction

Samples of cecal and rectal contents were frozen in liquid nitrogen and pulverized with an oscillating ball mill (MM400, Retsch Inc.). The homogenized samples were used for microbial DNA extraction according to the protocol described for [43] with modifications. Briefly, 0.25 g of sample was transferred to Lysing Matrix E Tube – Kit FastDNA Spin Kit for Soil (MP Biomedical, Inc.). For cell lysis, 1 mL RBB+C buffer was added in each sample, followed by homogenization in a FastPrep® FP120 instrument (MP Biomedical, Inc.). The precipitation of nucleic acids was obtained with the addition of a solution of ammonium acetate (10 M). The samples were incubated on ice for 30 min and then centrifuged at  $4^{\circ}\text{C}$  for 10 min at  $16,000\times g$ . The nucleic acids pellet was recovered and washed with 70% (v.v<sup>-1</sup>) ethanol, followed by drying at room temperature. The nucleic acid pellet was dissolved in 75  $\mu\text{L}$  of autoclaved ultrapure water. RNA was removed with the addition of DNase-free RNase (10 mg.mL<sup>-1</sup>). DNA purification was performed using PowerClean® DNA Clean-Up Kit (Mo Bio Laboratories) according to the manufacturer's protocol. Finally, electrophoresis using 0.8% agarose gel was used to separate the DNA fragments and to evaluate DNA quality. The DNA solution was stored at  $-20^{\circ}\text{C}$ .

### RNA extraction and mRNA enrichment

The samples homogenized with an oscillating ball mill were also used for RNA extraction. Briefly, 500 mg of sample was used for total RNA extraction with Trizol and FastRNA® Pro Green Kit (MP Biomedicals),

according to the manufacturer's instructions. The total RNA samples were treated with Ribo-Zero™ Magnetic Kit (Epicentre Biotechnologies) to remove ribosomal RNA (rRNA) from total RNA and enrichment of mRNA, followed by a 5 min incubation at 50 °C with magnetic beads to remove the hybridized rRNA molecules from the mRNA. Subsequently, the supernatant was purified using an 80 % ethanol solution and the resultant RNA were used for sequencing.

### **Microbial community structure and diversity analysis**

Capybara gut microbial community structure and diversity was investigated via high-throughput sequencing of 16S rRNA gene. The amplification of the 16S rRNA gene V4 region was performed using the 515F (5'-GTGCCAGCMGCCGCGGTAA) and 806R (GGACTACHVGGGTWTCTAAT) primers [44]. Sequencing was performed on an MiSeq Sequencing System (Illumina Inc.) with the V3 kit, 600 Cycles, in paired-end sequencing mode 2x300bp. The ZymoBIOMICS™ Microbial Community DNA Standard (D6305, Zymo Research) with eight phylogenetically distant bacterial strains (3 gram-negative and 5 gram-positive) and 2 yeasts, was included as a positive control to evaluate possible bias in libraries construction, sequencing, and bioinformatics analysis. For taxonomy analysis, paired-end reads were quality checked using FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and filtered using Trimmomatic v.0.36 [45] to remove adapters and low-quality reads, using the following parameters: ILLUMINACLIP:NexteraPE-PE.fa:2:30:10:8:true, LEADING:4, TRAILING:4, MINLEN:60 and SLIDINGWINDOW:4:16. Filtered paired-end reads were merged using fastq\_mergepairs function from Usearch v.10 package [46] (parameters: fastq\_maxee 0.5, fastq\_minovlen 50, and fastq\_minmergelen 250). Detailed number of reads per sample are summarized in Suppl. Table S13. Prior to OTUs clustering, primer sequences were removed, and singletons were discarded. Filtered amplicon reads were denoised (error-corrected) using the UPARSE unoise3 function (parameters: -minsize 8 and alpha 0.2), to likely recover true biological sequences. Prokaryotic taxonomy assignment was performed using syntax function, as implemented in Usearch v.10, using a the syntax\_cutoff parameter of 0.8 as threshold and the RDP database v16 [47]. Further analyses were performed using phyloseq v1.20 R package on R Studio.

### **Metagenome and metatranscriptome sequencing**

Metagenomic libraries were prepared using the Nextera Library Preparation kit (Illumina Inc.), while metatranscriptomic libraries were prepared using the TruSeq Stranded total RNA Library Prep Kit (Illumina Inc.). Libraries concentration were measured through quantitative qPCR using the KAPA Library Quantification Kit (Roche Inc.) and assayed for quality using the BioAnalyzer (Agilent Technologies). MG and MT libraries were paired-end sequenced in two runs (2 × 100 bp) on the Illumina HiSeq 2500 platform at the NGS sequencing facility at LNBR/CNPEM (Campinas, Brazil). The cecal and rectal gDNA were homogenized into a single sample and sequenced on a MinION sequencing device from Oxford Nanopore Technologies Inc. to obtain long reads. About 1 µg of ultra-long high molecular weight gDNA from the homogenized samples was require for library preparation. The library was prepared with the SQK-LSK109 Kit (Oxford Nanopore Technologies Inc.) following the manufacturer protocol. The MinION run was performed on a Flow cell R9 version, generating around 3Gb of long reads.

## Metagenome and metatranscriptome analysis

Metagenome and Metatranscriptome raw sequences were quality checked and trimmed as described above, MT reads were also analyzed using SortmeRNA v. 2.0 to remove rRNA reads, and then both MG and MT reads were taxonomically classified using Kaiju v. 1.7.4 with maximum number of mismatches allowed=5 and with the greedy mode [48]. For functional analysis, the MG trimmed reads were *de novo* co-assembled using IDBA\_UD (version 1.1.1) with the pre-correction parameter and k-mer size from 20 to 60 [49]. Assembly statistics are described in Suppl. Table S14. The assembled metagenome was binned using CONCOCT v.0.4.0 (parameters: -c 400, -k 4 -l 1000, -r 200 and -no\_cov\_normalization) [50] and MaxBin 2.0 (parameters: min\_contig\_length 1000, max\_iteration 50, prob\_threshold 0.9 and markerset 107) [51] to recover putative genomes from the metagenomic data. The binned genomes were dereplicated to remove redundancies using dRep v. 2.0.5 (parameters: -comp 80 -con 10 -str 100 and -p 10) and analyzed using CheckM v1.0.6 with the lineage\_wf workflow [52] to determine the completeness and contamination ratios of these genomes. Long-reads sequencing were used for MAGs scaffolding using SSPACE-long-reads v1.1 (parameters: -k 5, -a 0.7, -x 1, -m 50, -o 20 and -n 1000). Genomes with completeness smaller than 55% and more than 15% contamination rate were discarded. To assign taxonomy to the recovered genomes GTDB-tk tool v.1.4 was used with the release 95 of the GTDB database [53]. Gene prediction and annotation of both the recovered genomes and the co-assembly were performed using Prokka v.1.11 with the meta parameter [54], and annotation statistics are described in Suppl. Table S14. KEGG pathways and Kegg Orthologous (KOs) annotation were performed using KOFAM (e-value < 1e-5) [55] and Functional Ontology Assignments for Metagenomes (FOAM) database (e-value < 1e-5) [56]. CAZymes annotation was performed according to CAZy database pipelines [57]. Furthermore, MG and MT reads were mapped to the whole set of genes recovered from the co-assembled metagenome and the set of genes recovered from the MAGs using Kallisto v. 0.46.1 with quant function [58] to estimate the coverage/abundance of protein coding genes in cecal and rectal samples. Normalized abundance was estimated based on the count/number of reads per kilobase per million mapped reads expressed as TPM.

## Phylogenetic analysis and metabolic reconstruction

Phylogenetic analysis of the MAG57, reference *Bacteroidetes* type strains and *Prevotellaceae* uncultured genomes recovered from UBA project [38] was performed using concatenated 92 single copy core genes according to UBCG method [59]. CAZymes phylogenetic analysis was carried out using the catalytic domain of each family aligned with MAFFT [60], and using maximum likelihood methods employing the RAxML software [61] with 1,000 rapid bootstrap inferences and LG as the substitution model.

To perform the metabolic pathways reconstruction of each recovered MAG, their annotation obtained from KOFAM database were filtered to keep only the top 5 hits of each protein with e-value below the 1e-5 threshold. These filtered annotations were then supplied to the Annotation of Metabolite Origins (AMON) tool [62], which based on the KOs annotated in each MAGs predicts the putative metabolites that it can generate.

## NMR-based metabolomics

Approximately 30 mg of dry cecal and rectal contents, and 300  $\mu\text{L}$  of solution 2:1 (methanol: chloroform) were mixed and sonicated for 1 min (4 cycles of 15 sec with intervals of 10 sec) and placed at 4 °C for 15 min. Next, 300  $\mu\text{L}$  of solution 1:1 (methanol: ultrapure water) was added, followed by centrifugation at 16,000 g and 4 °C for 20 min. The supernatant was transferred to a new tube and were dried in CentriVap Solvent System (Labconco Corporation). Samples were diluted to 630  $\mu\text{L}$  by addition of  $\text{D}_2\text{O}$ , 70  $\mu\text{L}$  of sodium phosphate buffer (final concentration 0.1 M) containing dimethyl-silapentane-sulfonate (final concentration 0.5 mM) for NMR chemical shift reference and concentration calibration. The samples were filtrated in a syringe filter with a 0.22  $\mu\text{m}$  pore size hydrophilic polyethersulfone (PES) membrane. The final volume of filtrate ranged from 500 to 650  $\mu\text{L}$ .  $^1\text{H}$  NMR spectra of samples were acquired using a Varian Inova NMR spectrometer (Agilent Technologies Inc.) equipped with a 5 mm triple resonance cold probe and operating at a  $^1\text{H}$  resonance frequency of 599.84 MHz and constant temperature of 298 K (25 °C). A total of 1024 free induction decays were collected with 32-k data points over a spectral width of 16 ppm. A 1.5-s relaxation delay was incorporated between scans, during which a continual water presaturation radio frequency (RF) field was applied. Spectral phase and baseline corrections, as well as the identification and quantification of metabolites present in samples, were performed using Chenomx NMR Suite 7.6 software (Chenomx Inc.).

## Protein expression and purification

Protein expression and purification were conducted as reported in [63]. Briefly, *E. coli* BL21 strain was transformed with target genes subcloned into pET28a in frame to a 6xHis-Tag at the N-terminus. LB medium [0.5% (w.v<sup>-1</sup>) yeast extract, 1% (w.v<sup>-1</sup>) tryptone, 1% (v.v<sup>-1</sup>) sodium chloride] was employed for protein expression in the presence of specific antibiotics for transformant selection. Culture growths were conducted at 37 °C until O.D.<sub>600nm</sub> around 0.8 and then, expression was induced by the addition of 0.2 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) (Sigma Aldrich). Protein expression was conducted at 18 °C for 16 hrs and cells were harvested by centrifugation at 5000 x g.

Protein purifications were performed by nickel affinity chromatography followed by size exclusion chromatography. Briefly, pelleted cells were resuspended in saline-phosphate buffer (20 mM sodium phosphate, 500 mM NaCl, pH 7.5) with addition of 5 mM imidazole, 1 mM phenylmethylsulfonyl fluoride (PMSF), 5 mM benzamidine and 0.1 mg ml<sup>-1</sup> lysozyme. Cells were then disrupted by sonication and soluble protein lysates were applied to a 5-ml HiTrap Chelating HP column (GE Healthcare). Target proteins were eluted in an imidazole gradient up to 0.5 M. 6xHis-Tag was cleaved using 1% (w.w<sup>-1</sup>) trypsin (catalog no. T1426, Sigma Aldrich). Target proteins were further purified by size exclusion chromatography with a HiLoad 16/600 Superdex 75 pg column (GE Healthcare) equilibrated with 20 mM sodium phosphate, 150 mM NaCl, pH 7.5. Purified proteins were evaluated by dynamic light scattering (DLS) and samples with low polydispersity (<20%) were employed in biochemical and biophysical experiments.

## Enzyme assays

Purified enzymes were screened against polysaccharides, oligosaccharides and synthetic substrates as described in Suppl. Table S8. The saturation curves of the enzymes GH43\_12 ( $\alpha$ -L-arabinofuranosidase, 35 °C and pH 6.0), GH97 ( $\alpha$ -galactosidase, 35 °C and pH 7.0) and GHXXX ( $\beta$ -galactosidase, 45 °C and pH 7.5) were obtained using the synthetic substrates *p*-nitrophenyl- $\alpha$ -L-arabinofuranoside (pNP- $\alpha$ -LaraF), *p*-nitrophenyl- $\alpha$ -D-galactopyranoside (pNP- $\alpha$ -D-Gal) or *p*-nitrophenyl- $\beta$ -D-galactopyranoside (pNP- $\beta$ -D-Gal) (Sigma-Aldrich), respectively, at the optimal pH and temperature of each enzyme in McIlvaine buffer (70 mM). The saturation curves of the full-length CapGH10 enzyme (50 °C and pH 5.5) and of the isolated GH10 catalytic domain (55 °C and pH 5.5) were determined for both xylan and rye arabinoxylan substrates using the 3,5-dinitrosalicylic acid method [64] in McIlvaine buffer (70 mM). Kinetic data were calculated from initial velocities and expressed as mean  $\pm$  SD from three independent experiments ( $n = 3$ ). Binding capacity of the carbohydrate-binding module from the GH10 enzyme and its mutants were evaluated by affinity gel electrophoresis (AGE) according to [65].

## Small Angle X-ray Scattering (SAXS)

SAXS data of the CapGH10  $\beta$ -helix domain were collected at the SAXS1 beamline (Brazilian Synchrotron Light Laboratory, Campinas, Brazil) at protein concentration of 8.4 mg.mL<sup>-1</sup> in 20 mM Hepes buffer pH 7.5. Buffer scattering were recorded and subtracted from the protein scattering. SAXS patterns were integrated using Fit2D [66] and GNOM [67] was used to evaluate the pair-distance distribution functions  $p(r)$ . *Ab initio* molecular envelope was calculated from SAXS data with DAMMIN [68] and an averaged model was generated from several runs using DAMAVER [69]. The crystallographic coordinates were fitted into the SAXS low-resolution model using SUPCOMB [70].

## Crystallization, diffraction data collection and structure determination

Crystallization experiments of the CapGH10  $\beta$ -helix domain were carried by the sitting-drop vapor-diffusion method at 18 °C. Native crystals were grown in 20% (w.v<sup>-1</sup>) PEG 6000, 0.1 M sodium acetate (pH 5.0) and 0.2 M sodium chloride. SeMet crystals were obtained under the same condition with the addition of 20 mM betaine hydrochloride. For cryoprotection, crystals were soaked in the reservoir solution added with glycerol or PEG 400 (20% (w.v<sup>-1</sup>)) prior to flash cooling. Diffraction datasets of native and SeMet crystals were collected at the PROXIMA-2A and MX2 beamlines from SOLEIL (Gif-sur-Yvette Cedex, France) and LNLS (Brazilian Synchrotron Light Laboratory, Campinas, Brazil), respectively. Datasets were indexed, integrated, merged, and scaled using XDS package [71]. The structure was solved by single anomalous dispersion (SAD) using the programs SHELXC, SHELXD and SHELXE [72] for data preparation, anomalous scatters location and phase calculation, respectively. An initial model was built with the AutoBuild Wizard [73] from the Phenix package [74]. The structure was refined with the programs PHENIX.REFINE [75] and REFMAC5 [76], and the models were inspected and manually adjusted according to the computed  $\sigma_A$ -weighted ( $2F_o - F_c$ ) and ( $F_o - F_c$ ) electron density maps using COOT [77]. TLS groups were calculated by TLSMD [78] and applied during the crystallographic refinement. The final structure

was evaluated by MolProbity [79] and PDBRedo server [80]. Structure factors and atomic coordinates were deposited at the Protein Data Bank (PDB) under the accession codes 7JVI. Data collection and refinement statistics are summarized in Suppl. Table S11.

## List Of Abbreviations

AGE: Affinity gel electrophoresis

CC: Cluster of CAZymes

CE: Carbohydrate esterase

CAZy: Carbohydrate-Active enZymes database

CAZymes: Carbohydrate-Active enZymes

GH: Glycoside hydrolase

MAG: Metagenome-assembled genome

KO: Kegg Orthologous

MG – Metagenomics

MT – Metatranscriptomics

NMR – Nuclear Magnetic Resonance

OMV: Outer membrane vesicle

OUT: Operational Taxonomic Unit

PL: Polysaccharide lyase

*p*NP- $\beta$ -D-Gal: *p*-nitrophenyl- $\beta$ -D-galactopyranoside

PUL: Polysaccharide utilization locus

RMSF: Rocky Mountain spotted fever

SCFA: Short-chain fatty acid

SeMet: Selenomethionine

T9SS: Type IX secretion system

UBA: Uncultured Bacteria and Archaea

# Declarations

## Ethics approval and consent to participate

This study was carried out in accordance with the Animal Management Rule of the Brazilian Ministry of Environment (Sisbio 59826-1). Fresh Samples collected from the cecum and rectum from three euthanized female animals were used for microbiome analysis.

## Consent for publication

Not applicable

## Availability of data and material

All sequencing data for this study can be found under the BioProject ID PRJNA563062. The 16S, metagenomic and metatranscriptome reads for cecal and rectal samples are available at SRA under the accession numbers SRR11852069-SRR11852086, SRR11852046-SRR11852057 and SRR11852097-SRR11852108, respectively (Suppl. Table S15). The MAGs can be found at GenBank under the accession numbers JABUSA0000000000 - JABUVA0000000000. Structural data have been deposited in the Protein Data Bank (<https://www.rcsb.org/>) under accession code 7JVI (CapCBMXX). All other data generated or analyzed during this study are included in this published article (and its Supplementary information files). Data and code used for microbiome analyses are publicly available at [https://github.com/gpersinoti/capybara\\_microbiome](https://github.com/gpersinoti/capybara_microbiome).

## Competing interests

The authors declare that they have no competing interests.

## Funding

This research was supported by grants from Fundação de Amparo à Pesquisa do Estado de São Paulo (grant no. 2015/26982-0 to M.T.M. and postdoctoral fellowship 2016/19995-0 to M.A.B.M) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (grant no. 305013/2020-3 to M.T.M., 408600/2018-7 to G.F.P, 439195/2016-0 to L.C, 150552/2017-3, and 142332/2017-8 to M.P.M.). L.C, "Coordenação de Aperfeiçoamento de Pessoal de Nível Superior" (CAPES; Finance code 001)".

## Authors contributions

LC, GFP and MTM designed the study and wrote the paper.

GFP, LFM, NT, VL and BH performed the multi-omics data analyses.

LC, DAP, MPM and MC performed the 16S, metagenomics and metatranscriptomics experiments.

LC, MPM, MC, MND, RASP, WCG and CAS expressed and purified the enzymes and performed the functional characterization.

LC, MPM, MABM, WCG and MTM performed the structural analysis.

LC and MLS performed the metabolomics analysis.

## Acknowledgements

We acknowledge the Brazilian Synchrotron Light Laboratory (LNLS) for the provision of time on the MX2 and SAXS1 beamlines, the Brazilian Biosciences National Laboratory (LNBio) for the use of the crystallization (Robolab), NMR and spectroscopy facilities, and the Brazilian Biorenewables National Laboratory (LNBR) for the use of the characterization of macromolecules and next generation sequencing facilities. LNLS, LNBio and LNBR are operated by the Brazilian Center for Research in Energy and Materials for the Brazilian Ministry for Science, Technology, Innovations and Communications. We acknowledge SOLEIL for provision of synchrotron radiation facilities at PROXIMA-2A (proposal 20181915) and we would like to thank William Shepard and Martin Savko for assistance in using the beamline. We acknowledge the support from Alana H. S. Alvarenga in protein purification and solubilization assays.

## References

1. Flint HJ, Bayer EA, Rincon MT, Lamed R, White BA. Polysaccharide utilization by gut bacteria: Potential for new insights from genomic analysis. *Nat Rev Microbiol.* 2008;6:121–31.
2. Morrison M, Pope PB, Denman SE, McSweeney CS. Plant biomass degradation by gut microbiomes: more of the same or something new? *Curr Opin Biotechnol.* 2009;20:358–63.
3. White BA, Lamed R, Bayer EA, Flint HJ. Biomass utilization by gut microbiomes. *Annu Rev Microbiol.* 2014;68:279–96.
4. Kartzinel TR, Hsing JC, Musili PM, Brown BRP, Pringle RM. Covariation of diet and gut microbiome in African megafauna. *Proc Natl Acad Sci U S A.* 2019;116: 23588–93.
5. Krause DO, Denman SE, Mackie RI, Morrison M, Rae AL, Attwood GT, et al. Opportunities to improve fiber degradation in the rumen: Microbiology, ecology, and genomics. *FEMS Microbiol Rev.* 2003;27:663–93.
6. 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;331:463–67.
7. Rincon MT, Ding SY, McCrae SI, Martin JC, Aurilia V, Lamed R, et al. Novel organization and divergent dockerin specificities in the cellulosome system of *Ruminococcus flavefaciens*. *J Bacteriol.* 2003;185:703–13.
8. Burnet MC, Dohnalkova AC, Neumann AP, Lipton MS, Smith RD, Suen G, et al. Evaluating Models of Cellulose Degradation by *Fibrobacter succinogenes* S85. *PLoS One.* 2015;10:e0143809.

9. Demment MW, Van Soest PJ. A nutritional explanation for body-size patterns of ruminant and nonruminant herbivores. *Am Nat.* 1985;125:641–72.
10. Stevens CE, Hume ID. Contributions of microbes in vertebrate gastrointestinal tract to production and conservation of nutrients. *Physiol Rev.* 1998;78:393–427.
11. Sakaguchi E. Digestive strategies of small hindgut fermenters. *Anim Sci J.* 2003;74:327–37.
12. Kiani A, Clauss M, Ortmann S, Vendl C, Congdon ER, Herrera EA, et al. Digestive physiology of captive capybara (*Hydrochoerus hydrochaeris*). *Zoo Biol.* 2019;38:167–79.
13. Herrera EA. Capybara digestive adaptations. In: Moreira JR, Ferraz KMPMB, Herrera EA, Macdonald DW, editors. *Capybara: biology, use and conservation of an exceptional neotropical species*. Springer; 2013. p. 97–106.
14. Polo G, Mera Acosta C, Labruna MB, Ferreira F, Brockmann D. Hosts mobility and spatial spread of *Rickettsia rickettsii*. *PLOS Comput Biol.* 2018;14:e1006636.
15. Delogu F, Kunath BJ, Evans PN, Arntzen M, Hvidsten TR, Pope PB. Integration of absolute multi-omics reveals dynamic protein-to-RNA ratios and metabolic interplay within mixed-domain microbiomes. *Nat Commun.* 2020;11:1–12.
16. Hackstein JHP, van Alen TA. Fecal methanogens and vertebrate evolution. *Evolution.* 1996;50:559–72.
17. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol.* 2017;35:725–31.
18. Armstrong Z, Mewis K, Liu F, Morgan-Lang C, Scofield M, Durno E, et al. Metagenomics reveals functional synergy and novel polysaccharide utilization loci in the *Castor canadensis* fecal microbiome. *ISME J.* 2018;12:2757–69.
19. Barker CJ, Gillett A, Polkinghorne A, Timms P. Investigation of the koala (*Phascolarctos cinereus*) hindgut microbiome via 16S pyrosequencing. *Vet Microbiol.* 2013;167:554–64.
20. Morrison PK, Newbold CJ, Jones E, Worgan HJ, Grove-White DH, Dugdale AH, et al. The Equine Gastrointestinal Microbiome: Impacts of Age and Obesity. *Front Microbiol.* 2018;9:3017.
21. Pratama R, Schneider D, Böer T, Daniel R. First Insights Into Bacterial Gastrointestinal Tract Communities of the Eurasian Beaver (*Castor fiber*). *Front Microbiol.* 2019;10:1646.
22. Velasco-Galilea M, Piles M, Viñas M, Rafel O, González-Rodríguez O, Guivernau M, et al. Rabbit Microbiota Changes Throughout the Intestinal Tract. *Front Microbiol.* 2018;9:2144.
23. O’Donnell MM, Harris HMB, Ross RP, O’Toole PW. Core fecal microbiota of domesticated herbivorous ruminant, hindgut fermenters, and monogastric animals. *Microbiologyopen.* 2017;6:e00509.
24. Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, et al. Evolution of mammals and their gut microbes. *Science.* 2008;320:1647–51.
25. Li J, Zhong H, Ramayo-Caldas Y, Terrapon N, Lombard V, Potocki-Veronese G, et al. A catalog of microbial genes from the bovine rumen unveils a specialized and diverse biomass-degrading

- environment. *Gigascience*. 2020;9:1–15.
26. Hagen LH, Brooke CG, Shaw CA, Norbeck AD, Piao H, Arntzen M, et al. Proteome specialization of anaerobic fungi during ruminal degradation of recalcitrant plant fiber. *ISME J*. 2021;15: 421–34.
  27. Ransom-Jones E, Jones DL, McCarthy AJ, McDonald JE. The *Fibrobacteres*: An Important Phylum of Cellulose-Degrading Bacteria. *Microb Ecol*. 2012;63:267–81.
  28. Raut MP, Couto N, Karunakaran E, Biggs CA, Wright PC. Deciphering the unique cellulose degradation mechanism of the ruminal bacterium *Fibrobacter succinogenes* S85. *Sci Rep*. 2019;9:1–15.
  29. Arntzen M, Várnai A, Mackie RI, Eijsink VGH, Pope PB. Outer membrane vesicles from *Fibrobacter succinogenes* S85 contain an array of carbohydrate-active enzymes with versatile polysaccharide-degrading capacity. *Environ Microbiol*. 2017;19:2701–14.
  30. Tamura K, Hemsworth GR, Déjean G, Rogers TE, Pudlo NA, Urs K, et al. Molecular mechanism by which prominent human gut *Bacteroidetes* utilize mixed-linkage beta-glucans, major health-promoting cereal polysaccharides. *Cell Rep*. 2017;21:417–30.
  31. de Souza AP, Leite DCC, Pattathil S, Hahn MG, Buckeridge MS. Composition and structure of sugarcane cell wall polysaccharides: implications for second-generation bioethanol production. *BioEnergy Res*. 2013;6:564–79.
  32. Martens EC, Lowe EC, Chiang H, Pudlo NA, Wu M, McNulty NP, et al. Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol*. 2011; 9:e1001221.
  33. Oliveira K, Costa C, Bittar CMM, Santos VP dos, Oliveira VAB, Sá JC. Indigestible cellulose and lignin in determining feces production and apparent digestibility in horses. *Acta Sci Anim Sci*. 2012;34:267–72.
  34. Gomez A, Sharma AK, Grev A, Sheaffer C, Martinson K. The Horse Gut Microbiome Responds in a Highly Individualized Manner to Forage Lignification. *J Equine Vet Sci*. 2021;96:103306.
  35. Hungate RE. *The Rumen and its Microbes*. 1st ed. New York: Academic Press; 1966.
  36. Rémond D, Ortigues I, Jouany J-P. Energy substrates for the rumen epithelium. *Proc Nutr Soc*. 1995;54:95–105.
  37. Makki K, Deehan EC, Walter J, Bäckhed F. The Impact of Dietary Fiber on Gut Microbiota in Host Health and Disease. *Cell Host Microbe*. 2018;23:705–15.
  38. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol*. 2017;2:1533–42.
  39. Grépinet O, Chebrou MC, Béguin P. Purification of *Clostridium thermocellum* xylanase Z expressed in *Escherichia coli* and identification of the corresponding product in the culture medium of *C. thermocellum*. *J Bacteriol*. 1988;170:4576–81.
  40. Schubot FD, Kataeva IA, Blum DL, Shah AK, Ljungdahl LG, Rose JP, et al. Structural basis for the substrate specificity of the feruloyl esterase domain of the cellulosomal xylanase Z from *Clostridium thermocellum*. *Biochemistry*. 2001;40:12524–32.

41. Larsbrink J, Rogers TE, Hemsworth GR, McKee LS, Tauzin AS, Spadiut O, et al. A discrete genetic locus confers xyloglucan metabolism in select human gut *Bacteroidetes*. *Nature*. 2014;506:498–502.
42. Louis P, Hold GL, Flint HJ. The gut microbiota, bacterial metabolites and colorectal cancer. *Nat Rev Microbiol*. 2014;12:661–72.
43. Yu Z, Morrison M. Improved extraction of PCR-quality community DNA from digesta and fecal samples. *Biotechniques*. 2004;36:808–12.
44. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A*. 2011;108:4516–22.
45. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
46. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods*. 2013;10:996–8.
47. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, et al. Ribosomal Database Project: Data and tools for high throughput rRNA analysis. *Nucleic Acids Res*. 2014;42:D633-42.
48. Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat Commun*. 2016;7:11257.
49. Peng Y, Leung HCM, Yiu SM, Chin FYL. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*. 2012;28:1420–8.
50. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning metagenomic contigs by coverage and composition. *Nat Methods*. 2014;11:1144–6.
51. 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.
52. 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.
53. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*. 2019;36:1925–27.
54. Seemann T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30:2068–69.
55. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, et al. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics*. 2020;36:2251–52.
56. Prestat E, David MM, Hultman J, Taş N, Lamendella R, Dvornik J, et al. FOAM (Functional Ontology Assignments for Metagenomes): A Hidden Markov Model (HMM) database with environmental focus. *Nucleic Acids Res*. 2014;42:e145.
57. 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.

58. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol.* 2016;34:525–7.
59. Na S-II, Kim YO, Yoon S-HH, Ha S min, Baek I, Chun J. UBCG: Up-to-date bacterial core gene set and pipeline for phylogenomic tree reconstruction. *J Microbiol.* 2018;56:280–5.
60. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol.* 2013;30:772–80.
61. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30:1312–3.
62. Shaffer M, Thurimella K, Quinn K, Doenges K, Zhang X, Bokatzian S, et al. AMON: Annotation of metabolite origins via networks to integrate microbiome and metabolome data. *BMC Bioinformatics.* 2019;20:614.
63. Santos CR, Costa PACR, Vieira PS, Gonzalez SET, Correa TLR, Lima EA, et al. Structural insights into  $\beta$ -1,3-glucan cleavage by a glycoside hydrolase family. *Nat Chem Biol.* 2020;16:920–9.
64. Miller GL. Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. *Anal Chem.* 1959;31:426–8.
65. Mandelli F, De Morais MAB, De Lima EA, Oliveira L, Persinoti GF, Murakami MT. Spatially remote motifs cooperatively affect substrate preference of a ruminal GH26-type endo- $\beta$ -1,4-mannanase. *J Biol Chem.* 2020;295:5012–21.
66. Hammersley AP, Svensson SO, Hanfland M, Fitch AN, Häusermann D. Two-dimensional detector software: From real detector to idealised image or two-theta scan. *High Press Res.* 1996;14:235–48.
67. Svergun DI. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Crystallogr.* 1992;25:495–503.
68. Svergun DI. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J.* 1999;76:2879–86.
69. Volkov V V., Svergun DI. Uniqueness of ab initio shape determination in small-angle scattering. *J Appl Crystallogr.* 2003;36:860–4.
70. Kozin MB, Svergun DI. Automated matching of high- and low-resolution structural models. *J Appl Crystallogr.* 2001;34:33–41.
71. Kabsch W. XDS. *Acta Crystallogr Sect D Biol Crystallogr.* 2010;66:125–32.
72. Sheldrick GM. Experimental phasing with SHELXC/D/E: Combining chain tracing with density modification. *Acta Crystallogr Sect D Biol Crystallogr.* 2010;66:479–85.
73. Terwilliger TC, Grosse-Kunstleve RW, Afonine P V., Moriarty NW, Zwart PH, Hung LW, et al. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr Sect D Biol Crystallogr.* 2007;64:61–9.
74. Adams PD, Afonine P V., Bunkóczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr Sect D Biol Crystallogr.* 2010;66:213–21.

75. Afonine P V., Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr Sect D Biol Crystallogr.* 2012;68:352–67.

76. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr Sect D Biol Crystallogr.* 1997;53:240–55.

77. Emsley P, Cowtan K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr Sect D Biol Crystallogr.* 2004;60:2126–32.

78. Painter J, Merritt EA. TLSMD web server for the generation of multi-group TLS models. *J Appl Crystallogr.* 2006;39:109–11.

79. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr Sect D Biol Crystallogr.* 2010;66:12–21.

80. Joosten RP, Long F, Murshudov GN, Perrakis A. The PDB-REDO server for macromolecular structure model optimization. *IUCrJ* 2014;1:213–20.

## Tables

**Table 1: Kinetic parameters of CAZymes heterologously expressed in *E. coli* BL21.**

Protein ID	CAZy		T		K <sub>M</sub>	k <sub>cat</sub> (s <sup>-1</sup> )	k <sub>cat</sub> /K <sub>M</sub>
	family	Substrate	pH	(°C)			
09512	GH97	pNP-α-D-Gal	7.0	35	8.43 ± 0.57 (mM)	34.1 ± 0.98	4.05
09513 (full-length)					2.14 ± 0.44 (mg.mL <sup>-1</sup> )	127.7 ± 16.3	59.67
09513 (GH10 domain)	GH10	Rye arabinoxylan	5.5	55	1.93 ± 0.08 (mg.mL <sup>-1</sup> )	180.1 ± 5.3	93.31
					1.69 ± 0.08 (mg.mL <sup>-1</sup> )	160.6 ± 5.22	95.03
		Xylan	5.5	55	2.74 ± 0.29 (mM)	151.19 ± 6.21	55.18
09514	GH43_12	α-L-arabinofuranoside	6.5	35	0.57 ± 0.05 (mM)	17.6 ± 0.39	30.88
44807		pNP-β-D-Gal			1.19 ± 0.35 (mM)	29.85 ± 1.95	25.08
CBK67650.1 β-Gal Domain	GHXXX	pNP-β-D-Gal	7.5	45			

# Figures

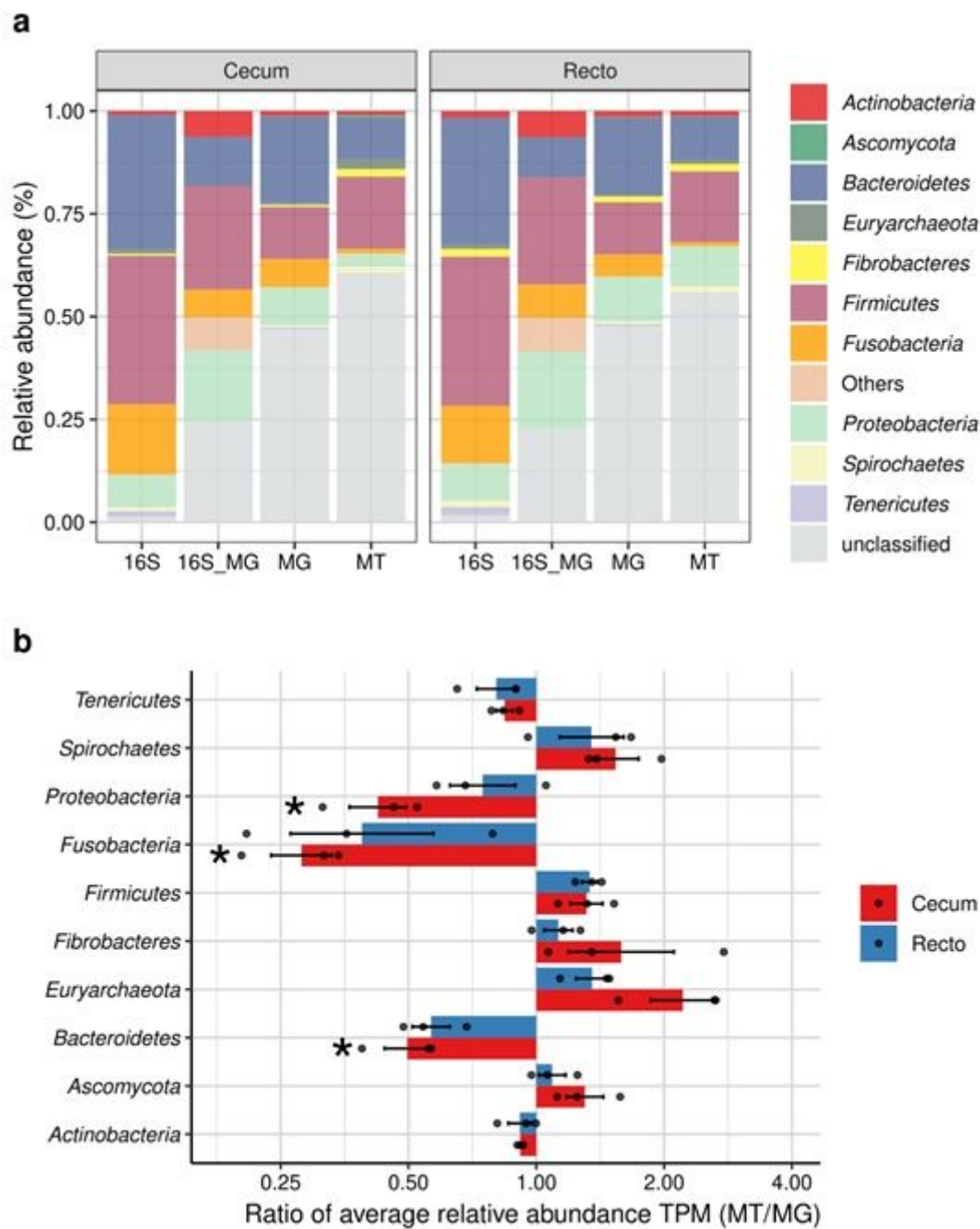
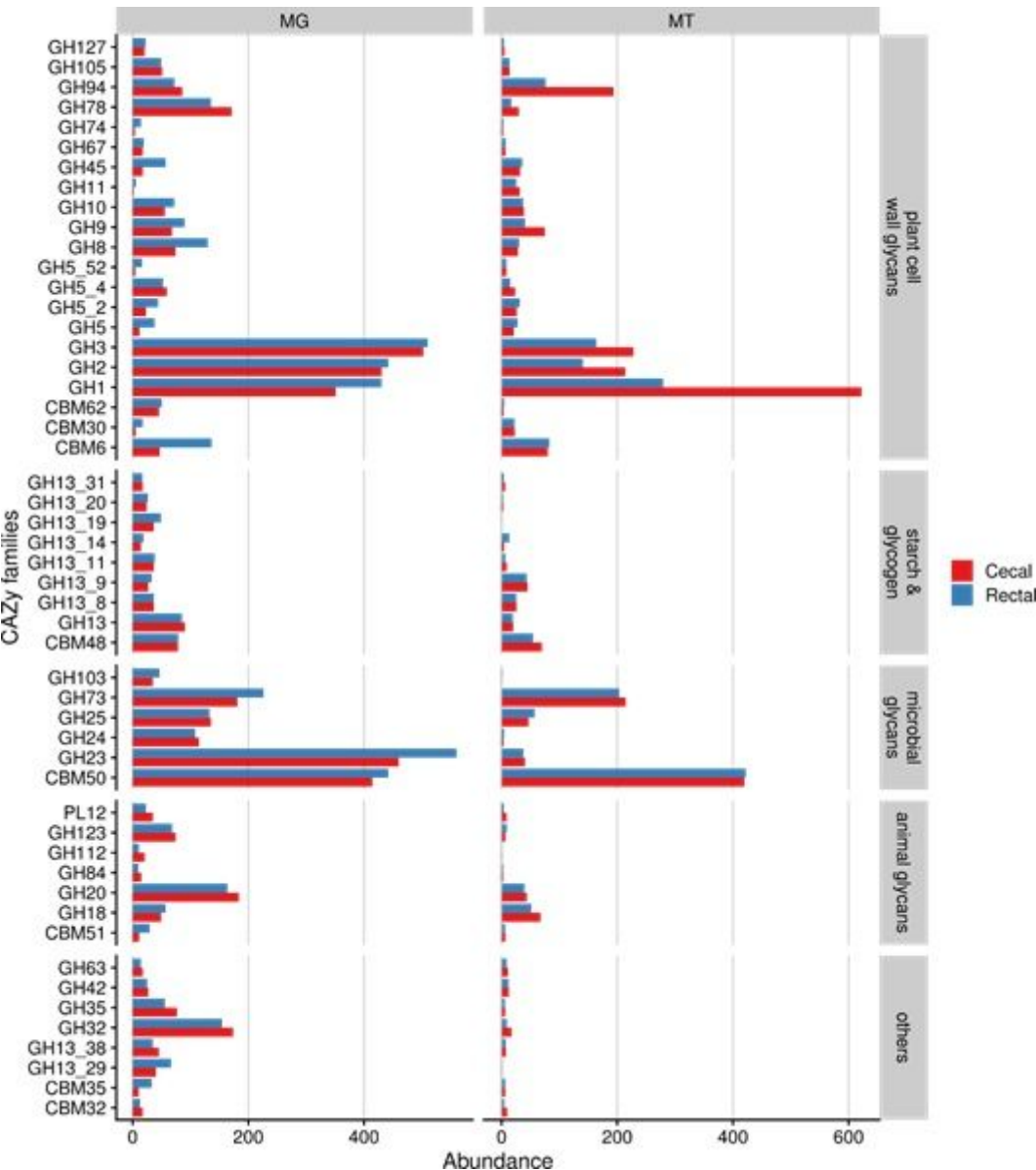


Figure 1

Microbial taxonomic composition of capybara gut microbiome. (a) Relative phyla abundance based on 16S rRNA gene target sequencing (16S), 16S rRNA recovered from metagenome (16S MG), whole metagenome (MG) and metatranscriptome (MT) reads. (b) Ratio of average relative phyla abundance (TPM) of metatranscriptome to metagenome (MT/MG), expressed in log<sub>2</sub> scale. Asterisks represent

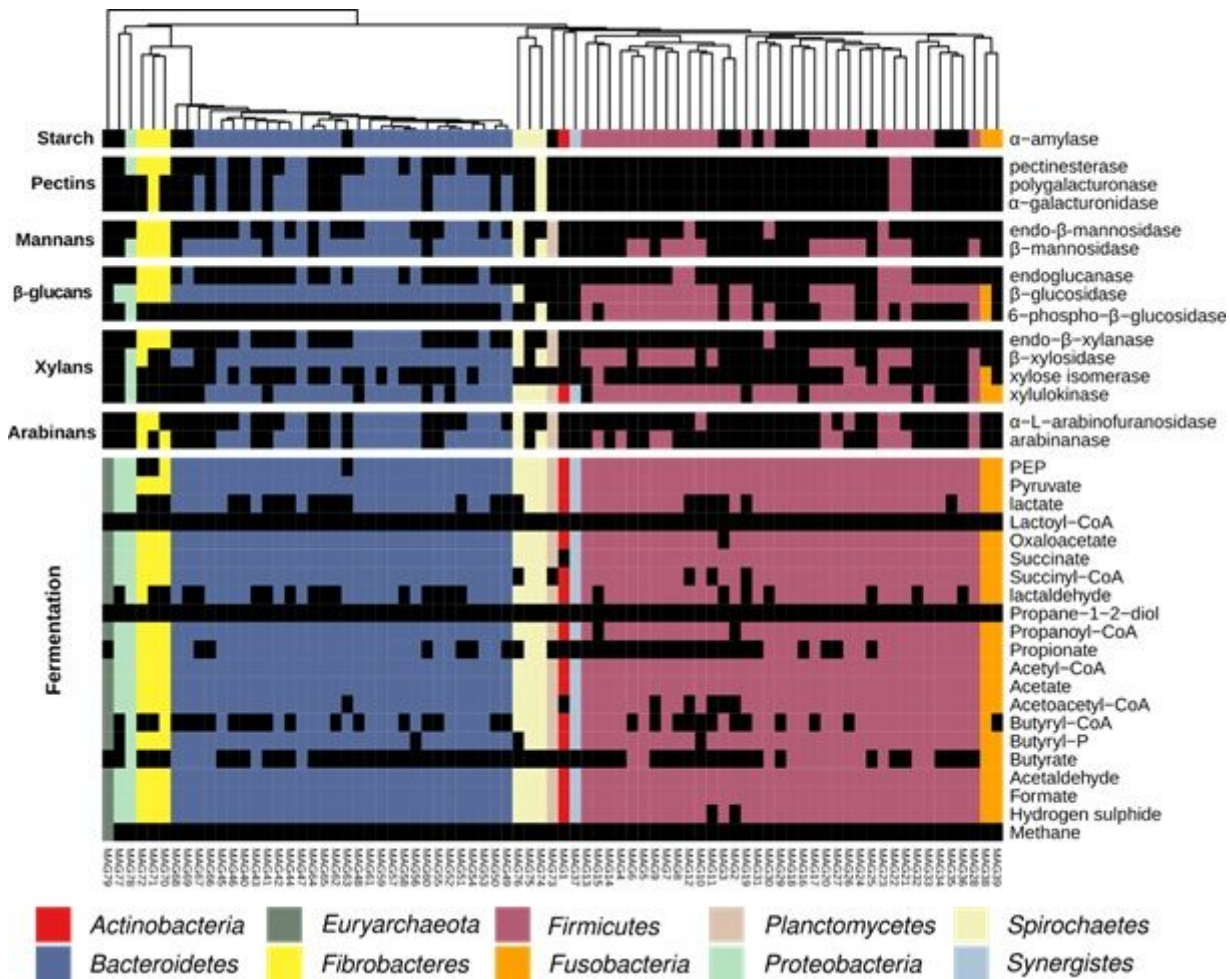
significantly different to MT/MG=1 (two-tailed t-test, FDR of 0.05 as threshold of significance level). The data represent the average of three independent experiments.



**Figure 2**

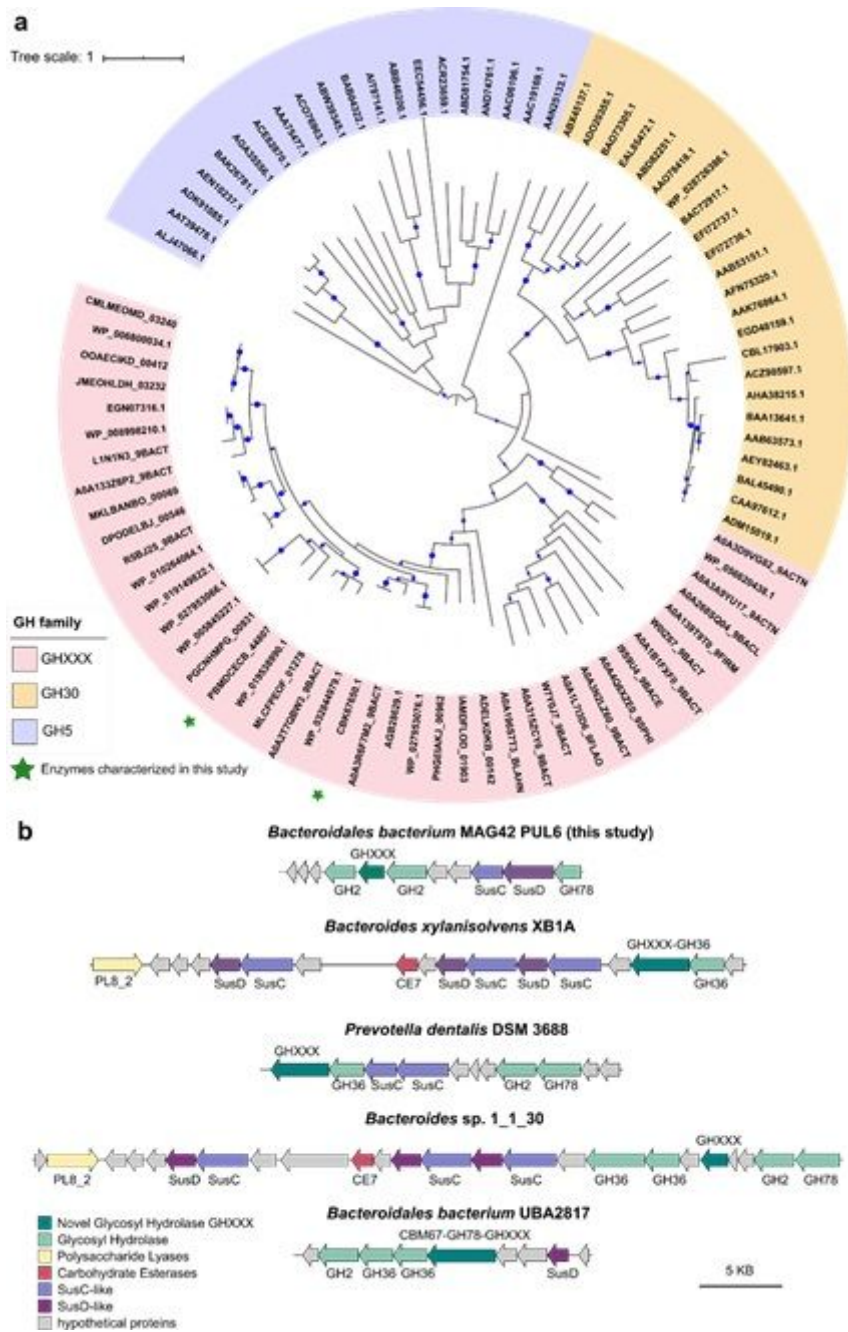
Functional annotation of Capybara metagenome co-assembly predicted genes according to CAZy database. Cumulative Abundance of main CAZy families predicted to act on distinct carbohydrates. Abundance is expressed as the cumulative TPM (Transcripts per Million). MG: Metagenome and MT: Metatranscriptome.





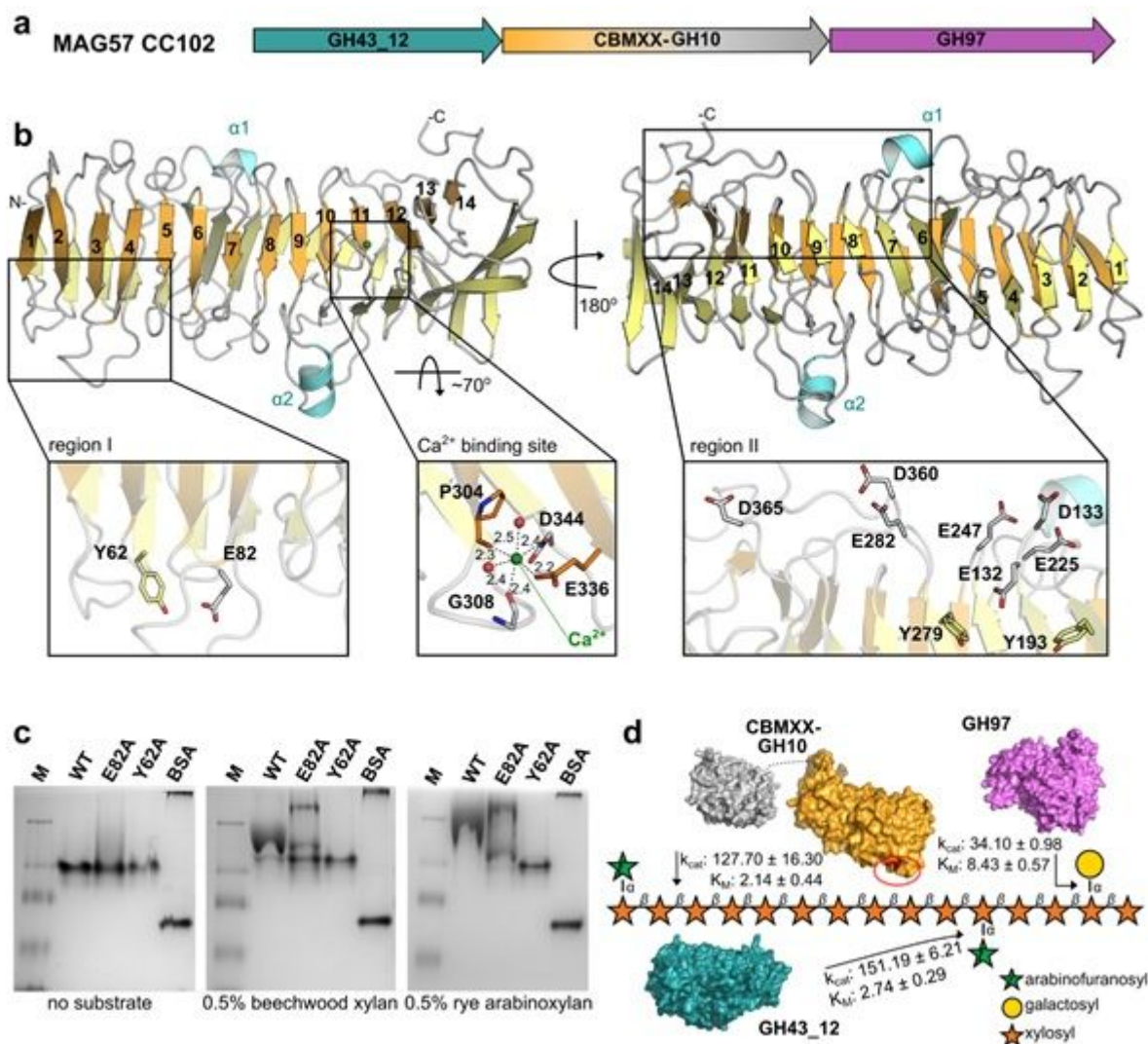
**Figure 4**

Metabolic reconstruction of 79 unique metagenome assembled genomes recovered from the capybara gut microbiome. Heatmap indicates the presence or absence of enzymatic systems related to plant polysaccharides degradation or metabolites production (listed on the left) in each metagenome-assembled genome (bottom) according to their set of genes. The presence of each enzyme/metabolite is denoted by a box colored by phylum taxonomy assignment, and black squares indicate the absence of the indicated metabolite/enzyme. Heatmap is clustered according to the phylogeny of the metagenome-assembled genomes.



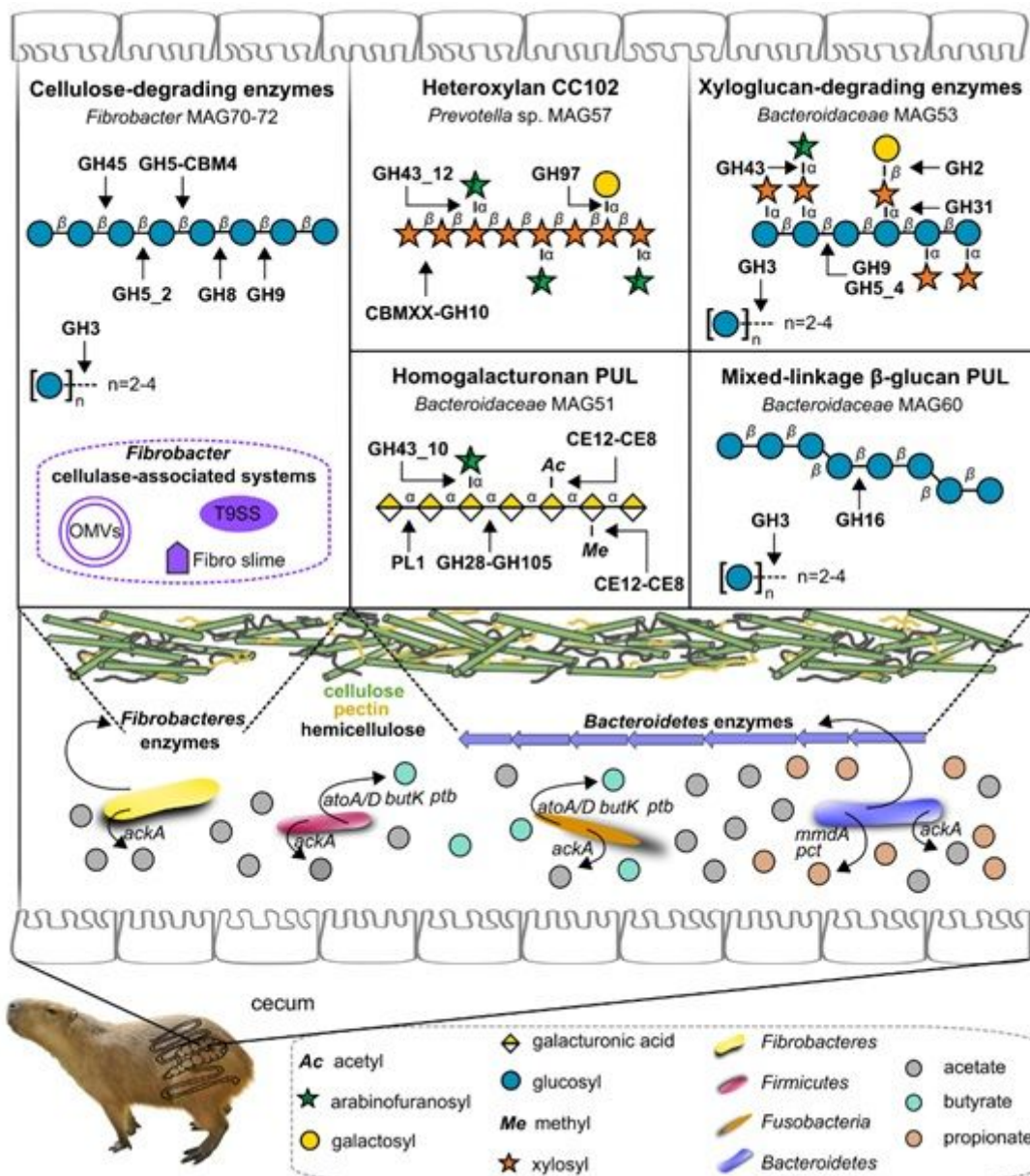
**Figure 5**

Phylogeny of the novel CAZy family GHXXX and its genomic context. (a) Maximum likelihood phylogenetic analysis of the GHXXX family (pink background) including characterized members from families GH30 (orange background) and GH5 (purple background). Nodes with bootstrap support values > 50 are indicated by the blue circles. Founding members of GHXXX family characterized in this study are denoted with a green star. (b) Genomic context of GHXXX-containing PULs identified in *Bacteroidetes* metagenome-assembled genomes.



**Figure 6**

Enzymatic system for heteroxylan degradation from *Prevotella* sp. MAG57. (a) Schematic representation of the CAZyme cluster (CC102) involved in heteroxylan breakdown. (b) Crystal structure of CapCBMXX indicating the  $\beta$ -helix fold consisting of 14 helical turns. The mutated residues from regions I and II, and the  $\text{Ca}^{2+}$  binding site are zoomed in (rectangles). (c) Affinity gel electrophoresis (AGE) of CapCBMXX (WT and mutants from region I) with xylan and arabinoxylan. Bovine serum albumin (BSA) was used as control (control) and M indicates the molecular weight markers. (d) Schematic representation of the modes of action of the enzymes CapCBMXX-GH10, CapGH97 and CapGH43\_12 on heteroxylans.  $k_{\text{cat}}$  values are expressed in  $\text{s}^{-1}$  and  $K_M$  in mM (for CapGH97 and CapGH43\_12) and  $\text{mg.mL}^{-1}$  for CapGH10. Protein surface representations were based on families scaffolds. The residues Tyr62 and Glu82, critical for xylan interaction, are highlighted (red) in the CapCBMXX surface (orange) and the approximated possible binding region (region I) is indicated (red circle).



**Figure 7**

Schematic representation of the capybara gut microbial community and enzymatic strategies involved in the deconstruction and conversion of typical dietary polysaccharides into short chain fatty acids. In the upper panel is highlighted the CAZymes and mechanisms accounted to the depolymerization of cellulose, main hemicelluloses and pectins found in grasses such as sugarcane. In the lower panel is shown the key phyla associated with hexoses and pentoses conversion into short-chain fatty acids. The upper and lower repeating drawings are standing for the capybara intestinal villi. Capybara image credits Tristan Tan/Shutterstock.com.

## Supplementary Files

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

- [SupplementaryFigures.pdf](#)
- [SupplementaryTables.xlsx](#)