



## Investigating Host Microbiota Relationships Through Functional Metagenomics

Elisabeth Laville, Josette Perrier, Nada Bejar, Marc Maresca, Jeremy Esque, Alexandra Tauzin, Emma Bouhajja, Marion Leclerc, Elodie Drula, Bernard Henrissat, et al.

### ► To cite this version:

Elisabeth Laville, Josette Perrier, Nada Bejar, Marc Maresca, Jeremy Esque, et al.. Investigating Host Microbiota Relationships Through Functional Metagenomics. *Frontiers in Microbiology*, 2019, 10, 12 p. 10.3389/fmicb.2019.01286 . hal-02315852

**HAL Id: hal-02315852**

**<https://hal.science/hal-02315852>**

Submitted on 14 Oct 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



# Investigating Host Microbiota Relationships Through Functional Metagenomics

Elisabeth Laville<sup>1\*</sup>, Josette Perrier<sup>2</sup>, Nada Bejar<sup>1</sup>, Marc Maresca<sup>2</sup>, Jeremy Esque<sup>1</sup>, Alexandra S. Tauzin<sup>1</sup>, Emna Bouhajja<sup>1</sup>, Marion Leclerc<sup>3</sup>, Elodie Drula<sup>4,5</sup>, Bernard Henrissat<sup>4,5</sup>, Stephane Berdah<sup>6,7</sup>, Eric Di Pasquale<sup>6,7</sup>, Patrick Robe<sup>8</sup> and Gabrielle Potocki-Veronese<sup>1</sup>

<sup>1</sup> INSA, INRA, CNRS, LISBP, Université de Toulouse, Toulouse, France, <sup>2</sup> iSm2, Centrale Marseille, CNRS, Aix-Marseille University, Marseille, France, <sup>3</sup> UMR1319, Micalis, Institut National de la Recherche Agronomique, Jouy-en-Josas, France, <sup>4</sup> CNRS, Architecture et Fonction des Macromolécules Biologiques, Aix-Marseille Université, Marseille, France, <sup>5</sup> USC 1408 AFMB, INRA, Marseille, France, <sup>6</sup> UMRT24 IFSTTAR, Laboratoire de Biomécanique Appliquée, Faculté de Médecine Secteur Nord, Aix-Marseille University, Marseille, France, <sup>7</sup> Inst Neurophysiopathol, INP, CNRS, Aix-Marseille Université, Marseille, France, <sup>8</sup> Givaudan SA, Toulouse, France

## OPEN ACCESS

### Edited by:

Vishal Singh,  
The University of Toledo,  
United States

### Reviewed by:

Alfonso Benítez-Páez,  
Instituto de Agroquímica y Tecnología  
de Alimentos (IATA), Spain  
Mirjam Perner,  
GEOMAR Helmholtz Centre for Ocean  
Research Kiel, Germany

### \*Correspondence:

Elisabeth Laville  
laville@insa-toulouse.fr

### Specialty section:

This article was submitted to  
Microbial Symbioses,  
a section of the journal  
Frontiers in Microbiology

**Received:** 11 December 2018

**Accepted:** 23 May 2019

**Published:** 07 June 2019

### Citation:

Laville E, Perrier J, Bejar N,  
Maresca M, Esque J, Tauzin AS,  
Bouhajja E, Leclerc M, Drula E,  
Henrissat B, Berdah S, Di Pasquale E,  
Robe P and Potocki-Veronese G  
(2019) Investigating Host Microbiota  
Relationships Through Functional  
Metagenomics.  
Front. Microbiol. 10:1286.  
doi: 10.3389/fmicb.2019.01286

The human Intestinal mucus is formed by glycoproteins, the O- and N-linked glycans which constitute a crucial source of carbon for commensal gut bacteria, especially when deprived of dietary glycans of plant origin. In recent years, a dozen carbohydrate-active enzymes from cultivated mucin degraders have been characterized. But yet, considering the fact that uncultured species predominate in the human gut microbiota, these biochemical data are far from exhaustive. In this study, we used functional metagenomics to identify new metabolic pathways in uncultured bacteria involved in harvesting mucin glycans. First, we performed a high-throughput screening of a fosmid metagenomic library constructed from the ileum mucosa microbiota using chromogenic substrates. The screening resulted in the isolation of 124 clones producing activities crucial in the degradation of human O- and N-glycans, namely sialidases,  $\beta$ -D-N-acetyl-glucosaminidase,  $\beta$ -D-N-acetyl-galactosaminidase, and/or  $\beta$ -D-mannosidase. Thirteen of these clones were selected based on their diversified functional profiles and were further analyzed on a secondary screening. This step consisted of lectin binding assays to demonstrate the ability of the clones to degrade human intestinal mucus. In total, the structural modification of several mucin motifs, sialylated mucin ones in particular, was evidenced for nine clones. Sequencing their metagenomic loci highlighted complex catabolic pathways involving the complementary functions of glycan sensing, transport, hydrolysis, deacetylation, and deamination, which were sometimes associated with amino acid metabolism machinery. These loci are assigned to several *Bacteroides* and *Feacalibacterium* species highly prevalent and abundant in the gut microbiome and explain the metabolic flexibility of gut bacteria feeding both on dietary and human glycans.

**Keywords:** functional metagenomics, carbohydrate-active enzymes, human intestinal mucin, human gut microbiota, lectin binding

## INTRODUCTION

The human intestinal tract harbors a dense commensal microbial community whose interactions with the host are of paramount importance for its health and well-being, particularly for the development of immunity and protection against invasion by pathogens. One of the best-known functions of the microbiota is its contribution to the digestion of dietary fibers, derived mainly from plant cell walls. Dietary fibers, consisting of polysaccharides of great structural diversity, represent the main energy source for the microbiota. The majority of gut bacterial species possess an enzymatic arsenal in the form of carbohydrate-active enzymes (CAZymes; Lombard et al., 2014) used to depolymerize these fibers. The products are further fermented to provide Short chain fatty acids (SCFAs) to the host. By contrast CAZymes produced by the human digestive system can only act on digestible starch and some other simple sugars such as sucrose or lactose (El Kaoutari et al., 2013). In addition to dietary fibers, intestinal microorganisms use glycans associated with human glycoproteins as a source of energy in particular mucin glycans that are secreted by the goblet cells of the intestinal epithelium. Mucin glycoproteins contain a large panel of mostly O-linked, and some N-linked, glycan structures (Podolsky, 1985; Robbe et al., 2004). Mucins represent a protective barrier for the intestine that is perpetually renewed. Mucus-colonizing bacteria produce both proteins to adhere to mucins and enzymes to degrade them (Tailford et al., 2015). The enzymatic machinery of these bacteria allows mucins to be recycled.

In recent years, growth assays on pig and human mucins, combined with pioneering genomic and transcriptomic studies, have shown that a number of pathogens and commensals (Tailford et al., 2015; Martens et al., 2018), including several *Bacteroides* species (Bäckhed et al., 2005; Sonnenburg et al., 2005; Martens et al., 2008; Marcobal et al., 2013), *Ruminococcus gnavus* (Owen et al., 2017) and the probiotic species *Akkermansia muciniphila* (Derrien et al., 2004), are involved in mucin degradation. This was especially described in cases of dietary fiber deprivation, when certain commensals shift their metabolism from dietary to host glycans (Desai et al., 2016). Furthermore, recent studies have also shown correlations between host diet, a reduction in mucus thickness, microbiota composition, and inflammatory bowel diseases (ulcerative colitis and Crohn's disease) and colorectal cancers (Pickard et al., 2014; Jakobsson et al., 2015; Desai et al., 2016). Despite these recent advances, only a dozen CAZymes (reviewed in Tailford et al., 2015) have been biochemically proven to degrade mucin oligosaccharides, and our understanding of mucin glycan-degrading pathways is almost exclusively restricted to cultured species. To date, only one enzyme has been characterized as involved in the degradation of human N-glycans by an uncultured gut bacterium (Ladevèze et al., 2013, 2015). Considering the fact that the majority of microbiota species are still uncultured, the studies so far conducted into the degradation of human glycans by gut bacteria are thus far from exhaustive.

In this study, we used functional metagenomics to identify mucin-degrading pathways from the uncultured fraction of the human gut microbiota. A two-step activity-based screening

strategy was applied to search for mucin catabolic pathways in a mucosal ileal metagenomic library, and to demonstrate their involvement in the structural modification of human colon mucus. Genomic and metagenomic data were integrated to assess the abundance and prevalence of these mucin glycan utilization loci in the human gut microbiome, and to reveal synteny with genomic loci from reference gut bacteria, which had not previously been proven to be involved in these host–microbiota relationships.

## MATERIALS AND METHODS

### Metagenomic DNA Sampling and Library Construction

Bacterial metagenomic DNA was obtained from the distal ileum sample of a 50–55-year-old patient undergoing colonoscopy and surgery for a suspected cancer of the lower colon after patient had been subjected to a cleansing preparation. The patient did not receive any antibiotics or other drugs in the 6 months before sampling. The sampling protocol was approved by the local ethics committee: the Comité de Protection des Personnes Sud Est V (Ref.: 07-CHUG-21; B. Habozit, J. L. Crakowski, J. Juge, J. Grunwald, E. Svhan, and E. Fontaine on the committee). A segment of 2 cm<sup>2</sup> was obtained from a healthy zone, before being immediately frozen and kept at –80°C until processing. The ileal mucosa was scraped, an enriched bacterial fraction was recovered using the method described in Courtois et al. (2003), and the metagenomic DNA was extracted as described in Tasse et al. (2010). Fragments of between 30 and 40 kb in size were isolated and cloned into the pCC1FOS fosmid (Epicenter Technologies). EPI100 *E. coli* cells were then transfected to obtain a library of 20,000 clones. Recombinant clones were transferred to 384-well microtiter plates containing a Luria-Bertani (LB) medium, supplemented with 12.5 mg/L chloramphenicol and 8% (w/v) glycerol. They were grown for a period of 22 h at 37°C and then frozen and conserved at –80°C. All other culture media mentioned in this study contained 12.5 mg/L chloramphenicol.

### Metagenomic Library Screening

The library was gridded using an automated microplate gridded (K2, KBiosystem, Basildon, United Kingdom) on 22 cm × 22 cm trays containing a solid agar medium supplemented with 12.5 mg/L chloramphenicol and with chromogenic substrates at a final concentration of 60 µg/mL (w/v). The chromogenic substrates used – 5-Bromo-4-chloro-3-indolyl α-L-fucopyranoside (X-α-L-Fuc), 5-Bromo-4-chloro-3-indolyl 2-acetamido-2-deoxy-β-D-glucopyranoside (X-β-D-GlcNAc), 5-Bromo-4-chloro-3-indolyl N-acetyl-α-D-neuraminic acid (X-α-D-Neu5Ac), 5-Bromo-4-chloro-3-indolyl 2-acetamido-2-deoxy-β-D-galactopyranoside (X-β-D-GalNAc), 5-Bromo-4-chloro-3-indolyl α-D-mannopyranoside (X-α-D-Man), and 5-Bromo-4-chloro-3-indolyl β-D-mannopyranoside (X-β-D-Man) – were purchased from Carbosynth<sup>1</sup>. The plates were incubated at 37°C for periods

<sup>1</sup><https://www.carbosynth.com>

of between 1 day and 2 weeks, depending on the time needed to observe the blue-colored clones.

## Lectin Binding Experiments

On the basis of their multiple activities on the chromogenic X-substrates from the primary screening, a selection of hit clones were tested for their ability to degrade human colon mucus using lectin binding assays. The hit clones were grown at 37°C in 400 mL LB medium, with orbital shaking at 120 rpm. After 16 h, cells were harvested using centrifugation for 5 min at 5,000 rpm, before being re-suspended and concentrated in an activity buffer (RPMI medium from Sigma) to obtain a final OD<sub>600</sub> nm of 80. Cell lysis was carried out using sonication. Cell debris were centrifuged at 13,000 rpm for 10 min and cytoplasmic extracts were filtered using a 0.20 µm Minisart RC4 syringe filter. An *E. coli* EPI100 clone containing the pCC1FOS fosmid without a metagenomic DNA fragment was used as a negative control. The human colon mucus was isolated from three patients of both genders, 31–74 years of age and all of European origin as previously described (Ajandouz et al., 2016). The sampling procedures were approved by the French Ethics Committee (CODECOH no. DC-2011-1319). Samples were taken from a macroscopically unaffected area as identified by the surgeon. After resection, the specimens were placed in an ice-cold DMEM solution supplemented with antibiotics. After longitudinal opening of the intestine, the surface of the mucosa was scraped in PBS to collect intestinal mucus. The mucus from the three patients was pooled and stored at –80°C until used. A lectin binding assay was used to test the effect of bacterial enzymes expressed by the positive clones on human mucins. Briefly, 96-well plates (F-Bottom, Nunc Maxisorp®) were coated overnight at 4°C with 100 µL of human mucins (5 µg/mL) diluted in a pH 9.6 bicarbonate/carbonate coating buffer (100 mM). After two washes with 200 µL phosphate saline buffer (pH 7.4) supplemented with Tween20 0.02% (v/v; PBS-T), the plates were incubated at 37°C with 100 µL of the bacterial supernatant (dilution: 1/100) for 5 h. The wells were washed twice with 200 µL PBS-T before being saturated at room temperature for 1 h with 200 µL PBS containing 1% BSA (PBS-B). Finally, the wells were washed twice with 200 µL PBS-Tween, and 100 µL of fluorescein-labeled lectin solution (5 µg/mL) in PBS-B were added. The fluorescein-conjugated lectins used (all from Vector Laboratories, France) were: Concanavalin A (Con A; selective of Man $\alpha$ 2/6/3-man $\beta$ 4-Glc(Nac)-R >  $\alpha$ -Man >  $\alpha$ -Glc >  $\alpha$ GlcNac), *Sambucus nigra* (soybean) agglutinin (SNA; selective of Neu5Ac $\alpha$ 1/6 Gal(Nac)-R) and *Triticum vulgaris* (wheat germ) agglutinin (WGA; selective of Gal $\beta$ 4-GlcNac $\beta$ 6/3-Gal $\beta$ 4-R > GlcNac $\beta$ -R > Neu5Ac $\alpha$ 3/6/8-R). After 1 h incubation at 37°C, the wells were washed and the fluorescence measured at an excitation/emission of 490 nm using a BioTek Synergy HTX Multi-Mode Microplate Reader (BioTek, France). Each binding experiment for a given lectin was performed in triplicate, using the same microplate for all hit clones including three control clones (*E. coli* host strain transformed with the empty vector Epi100).

The mean fluorescence value for the three control clones, corresponding to 100% binding, was used to calculate the

percentage of the lectin binding for each clone, including the three control clones. The variance within biological replicates of each clone was first assessed using the *F*-test against control clones. Subsequently, *p*-values were calculated using the *t*-test to evaluate the statistical significance of differences between a given clone and the controls. *P*-values lower than 0.05 were considered statistically significant.

## Metagenomic Sequence Analysis

The fosmid DNA of the clone hits was extracted using the NucleoBond Xtra Midi kit from Macherey-Nagel (France). Fosmids were then sequenced using the Ion Torrent S5 System at the GeT-Biopuces Platform (Toulouse, France). Read assembly was performed using Masurca<sup>2</sup>. The assembled contigs were cleaned from the pCC1FOS vector sequence using Crossmatch<sup>3</sup>. ORF detection and functional annotation was performed using the RAST annotation server<sup>4</sup> (Aziz et al., 2008). CAZyme encoding genes were identified by BLAST analysis of the predicted ORFs against the full-length sequences of glycoside hydrolases (GH), polysaccharide lyases (PL), carbohydrate esterases (CE), carbohydrate-binding modules (CBM), and glycosyltransferases (GT) included in the CAZy database<sup>5</sup>, using a cut-off *E*-value of 7·10<sup>-6</sup>. Sequences that aligned over their entire length with a sequence in the database with >50% identity were directly assigned to the same family as the subject sequence. The remaining ones were subjected in parallel to (i) a BLAST search against a library built with partial sequences corresponding to individual GH, PL, CE, CBM, and GT modules and (ii) a HMMER2 search using Hidden Markov models (HMM2) built for each CAZy module family, allowing a view of CAZyme modularity (Lombard et al., 2014). The contigs' entire nucleotide sequences were blasted against the non-redundant database of the NCBI and against each other to examine their divergence from the reference strain genomes, and the redundancy between clone hits. Taxonomic assignment of the metagenomic sequences was determined using the PhyloPythias program<sup>6</sup> (model type: Generic 2013–800 Genera). The same results were obtained with a minimum slice at 3% and 50%. The presence of signal peptide cleavage sites in ORF amino acid sequences were predicted using the SignalP server<sup>7</sup> (Nielsen, 2017).

## Prevalence Analyses

The homolog sequences of the translated ORFs were searched for in the translated catalog of 9.9 million reference genes using BLASTP (*E*-value = 0, identity ≥ 90%). This catalog consisted of the gut metagenomic sequences of 1,267 subjects from three continents (United States, China and Europe): 139 US HMP samples, 760 European fecal samples from the MetaHIT project and 368 Chinese fecal samples (Li et al., 2014). The microbial gene richness in the human gut was assessed by recovering

<sup>2</sup><http://www.genome.umd.edu/masurca.html>

<sup>3</sup><http://bozeman.mbt.washington.edu/phredphrapconsed.html>

<sup>4</sup><http://rast.nmpdr.org/>

<sup>5</sup><http://www.cazy.org>

<sup>6</sup><http://phylopythias.bifo.helmholtz-hzi.de>

<sup>7</sup><http://www.cbs.dtu.dk/services/SignalP/>



the occurrence frequency data of homologous sequences of the catalog from the 9.9 million gene frequency table in the 1,267 subjects<sup>8</sup>. The frequency values have no unit. They are normalized to account for sampling and sequencing biases generated by the diverse origins of the cohorts constituting the catalog.

## Data Deposition

The datasets generated in the course of this study are available in the repository of the DDBJ/EMBL/GenBank Nucleotide Sequence Database under accession numbers LR131274-LR131286<sup>9</sup>.

## RESULTS AND DISCUSSION

### Screening the Human Gut Metagenome for Mucin Glycan-Degrading Activities

The 20,000 clones of the *E. coli* fosmid library constructed from the ileum mucosa microbiota, covering in total 0.7 Gb of metagenomic sequence, were screened for nearly all the  $\alpha$ - and  $\beta$ -glycosidase activities required for the degradation of N- and O-glycans. Firstly,  $\alpha$ -L-fucosidases,  $\beta$ -D-N-acetyl-glucosaminidases,  $\beta$ -D-N-acetyl-galactosaminidases,  $\alpha$ -D-neuraminic-acid hydrolases,  $\alpha$ -D- and  $\beta$ -D-mannosidases were searched for using chromogenic reagents (5-Bromo-4-chloro-3-indoxyl-glycosyl, known as X-glycosyl substrates) in a rich solid medium on which the metagenomic clones had been gridded. It was not possible to screen for  $\beta$ -D-galactosidases using this approach due to the high background activity of the *E. coli* host.

These 120,000 assays allowed 158 validated activities to be identified, with this corresponding to 124 positive clones, since 40 clones produced several of the screened activities (Table 1). Strikingly, no positive clone was found on the fucosyl substrate, despite  $\alpha$ -linked-L-fucosyl residues frequently being found at the non-reducing end of O-linked oligosaccharides, and  $\alpha$ -fucosidase activity being widespread in gut bacteria whatever their taxa (Katayama et al., 2005; Tailford et al., 2015). This may be due to the fact that  $\alpha$ -fucosidases are known to be inactive on this kind of artificial substrate (Katayama et al., 2005). The hit yields varied from 0.1% (for X- $\alpha$ -D-Man) to 3.3% (for X- $\beta$ -D-GlcNac) depending on the screening substrate, which is on average 2.2 times higher than the hit rate (0.4–1.3%) obtained when screening the same library for dietary fiber hydrolytic activities (Cecchini et al., 2013).

### Assessment of Human Intestinal Mucus Degradation

Since chromogenic substrates do not represent the structural complexity of mucin oligosaccharides, we then confirmed the activity of the hit clones on the physiological substrate targeted by mucus-degrading bacteria. From the 124 positive clones, we selected 13 to be tested on human intestinal

mucus and/or sequenced. The clones chosen were those: (i) producing complementary glycosidase activities required to break down complex mucin oligosaccharide structures thanks to the expression of several CAZy-encoding genes clustering in the same metagenomic loci, or thanks to the production of a highly promiscuous CAZyme; and (ii) producing the highest levels of activity, defined as those with the quickest response to chromogenic testing.

The mucus-degrading ability of the metagenomic clones was assessed using the ELISA assay technique, which involves the modification of mucin-specific lectin binding if the epitopes are affected by CAZyme activity. The mucus was incubated with the hit clones' cytoplasmic extracts. After a washing step, it was then incubated with three fluorescent lectins specific to different N- and O-glycan motifs (Table 2). A comparison of lectin binding percentages with those obtained for reference clones (*E. coli* host strain transformed with the empty vector Epi100) is provided in Figure 1 and Table 2. We observed considerable variability in the binding inhibition values, likely due to the heterogeneousness of the glycan structures in the human mucus samples. Nevertheless, binding modification was significant for several clone–lectin couples. Overall, most of the

**TABLE 1 |** Results of primary screening.

Substrates	Number of positive clones – Hit rate
X- $\beta$ -D-GalNac	47 – 2,3%
X- $\beta$ -D-GlcNac	67 – 3,3%
X- $\alpha$ -D-Man	2 – 0,1%
X- $\alpha$ -D-Neu5Ac	27 – 1,3%
X- $\beta$ -D-Man	17 – 0,8%
X- $\alpha$ -L-Fuc	0
Single activities	Number of positive clones
X- $\beta$ -D-GalNac	15
X- $\beta$ -D-GlcNac	29
X- $\alpha$ -D-Man	1
X- $\alpha$ -D-Neu5Ac	20
X- $\beta$ -D-Man	9
Multiple activities	Number of positive clones
X- $\beta$ -D-Man + X- $\alpha$ -D-Neu5Ac	2
X- $\beta$ -D-GlcNac + X- $\alpha$ -D-Neu5Ac	1
X- $\beta$ -D-GlcNac + X- $\beta$ -D-Man	2
X- $\beta$ -D-GlcNac + X- $\beta$ -D-GalNac	27
X- $\beta$ -D-GlcNac + X- $\alpha$ -D-Man	1
X- $\beta$ -D-GlcNac + X- $\beta$ -D-GalNac + X- $\beta$ -D-Man	1
X- $\beta$ -D-GlcNac + X- $\beta$ -D-GalNac + X- $\alpha$ -D-Neu5Ac	3
X- $\beta$ -D-GlcNac + X- $\alpha$ -D-Man + X- $\beta$ -D-Man	2
X- $\beta$ -D-GlcNac + X- $\beta$ -D-GalNac + X- $\alpha$ -D-Neu5Ac + X- $\beta$ -D-Man	1

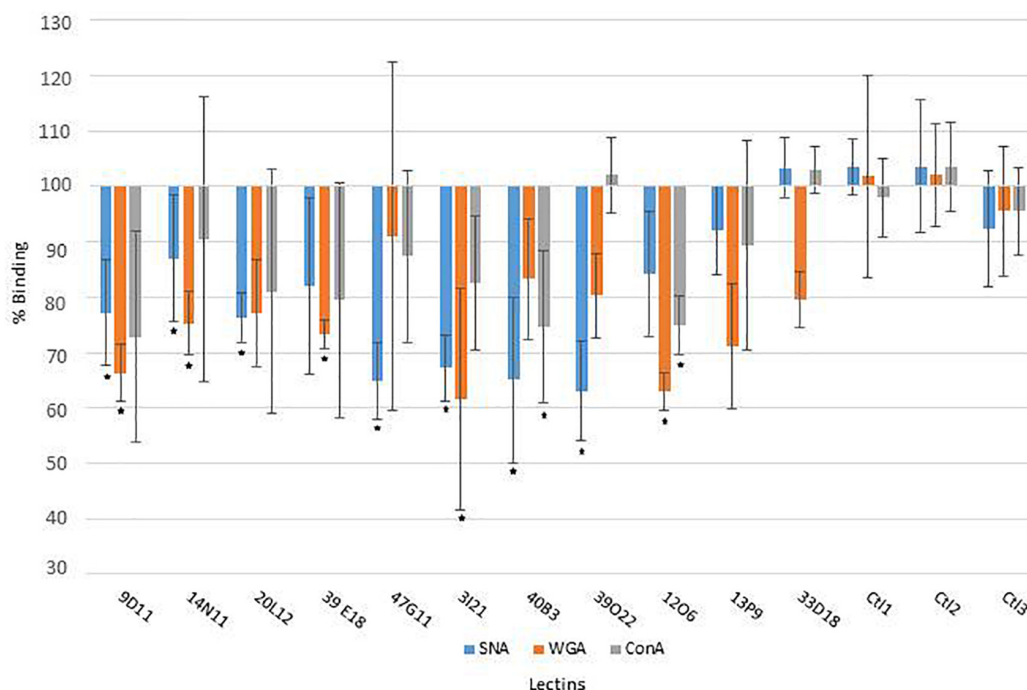
<sup>8</sup><http://meta.genomics.cn/meta/dataTools>

<sup>9</sup><https://www.ncbi.nlm.nih.gov/genbank/>

TABLE 2 | Functional profile and CAZyme-encoding gene content of the hit clones.

Clone identifier	CAZy families and their substrate target inferred from the activities listed in the CAZy database										Activity on X- substrates (GH family inferred as responsible for the clone phenotype)		Lectins and their binding specificities (GH family inferred as responsible for the clone phenotype)			
	$\beta$ -GlcNAc	$\beta$ -GalNAc	$\alpha$ -Neu5Ac	$\beta$ -Gal	$\alpha$ -Gal(Nac)	$\beta$ -Man	$\alpha$ -Man	$\alpha$ -Fuc	X- $\beta$ -GlcNAc	X- $\beta$ -GalNAc	X- $\alpha$ -Neu5Ac	X- $\beta$ -Man	WGA	SNA	ConA	
9D11				GH1	GH1	GH1			X			X(GH1)	Gal $\beta$ 4-GlcNAc $\beta$ 6/3-Gal $\beta$ 4-R	Neu5Ac $\alpha$ 1/6Gal(NAc)-R	Man $\alpha$ 2/6/3-man $\beta$ 4-Glc(Nac)- $\alpha$ GlcNAc	
14N11	GH20	GH20		GH2		GH2	GH2		X(GH20)	X(GH20)	X	X(GH2)	X(GH1)			
20L12	GH20	GH20	GH33	GH2	GH27	GH2			X(GH20)	X(GH20)	X(GH33)					
39E18	GH20	GH20	GH33	GH2	GH27	GH2			X(GH20)							
47G11	GH20	GH20	GH33	GH2		GH2	GH2	GH29			X(GH33)					
3I21	GH20	GH20	GH33					GH29	X(GH20)	X(GH20)	X(GH33)		X(GH20, GH33)	X(GH33)		
40B3			GH33								X(GH33)					
41E6	GH3	GH3	GH33								X(GH33)					
39O22			GH33								X(GH33)					
12O6	GH20	GH20		GH2		GH2			X(GH20)	X(GH20)						
47C24	GH20	GH20		GH2		GH2			X(GH20)	X(GH20)						
13P9				GH2		GH2						X(GH2)				
33D18				GH2		GH2						X(GH2)				

X, activity on X-substrates; \\_, significant decrease of lectin binding to colon mucus after incubation with the hit clones, compared to reference (*E. coli* host strain transformed with the empty vector Epi100); nt, non-tested; -, no significant activity/no alteration in binding.



**FIGURE 1 |** Lectin binding on human colon mucus after incubation with the clone extracts, compared to reference clones (*E. coli* host strain transformed with the empty vector Epi100; \*, significant modification of lectin binding).

selected clones, except 13P9 and 33D18, induced a decrease in binding of at least one lectin. The effects were particularly clear for the clones producing activities that target the  $\beta$ -D-GalNAc,  $\beta$ -D-GlcNAc, and  $\alpha$ -D-Neu5Ac residues usually found at the terminal extremities of human mucin glycans, which are the most accessible to exo-acting glycosidases. In most cases, the hit clones produced the activities required to degrade the glycan structures targeted by the specific lectins, thus affecting lectin binding.

With respect to the WGA lectin, which is specific to terminal Gal $\beta$ 4-GlcNAc $\beta$ 6/3-Gal $\beta$ 4-R > GlcNAc $\beta$ -R, and to a lesser extent, Neu5Ac $\alpha$ 3/6/8-R motifs, five  $\beta$ -D-N-acetyl-glucosaminidase-producing clones (9D11, 14N11, 39E18, 3I21, and 12O6) induced less lectin binding than was obtained for the reference clones. The terminal Gal $\beta$ 4 residues were likely removed by the *E. coli*  $\beta$ -galactosidase activity, facilitating access to the  $\beta$ -D-N-acetyl-glucosaminidases produced by the hit clones. Binding of the SNA lectin, which is specific to terminal Neu5Ac $\alpha$ 1/6 Gal(NAc)-R motifs, was reduced after the action of five clones active on X- $\alpha$ -D-Neu5Ac (20L12, 47G11, 3I21, 40B3, and 39O22) and, surprisingly, clone 9D11, which is not active on X- $\alpha$ -D-Neu5Ac. Also surprisingly, ConA lectin binding, which is specific to Man $\alpha$ 2/6/3-man $\beta$ 4-Glc(Nac)-R >  $\alpha$ -Man >  $\alpha$ -Glc >  $\alpha$ GlcNAc motifs, was inhibited after the action of two clones (40B3 and 12O6) which were not detected as being active on X- $\alpha$ -Man. We cannot exclude the possibility that these clones do indeed produce the enzymes required to break down the glycan motifs targeted by ConA, but that their activity cannot be detected on the corresponding X-glycosyl substrates due to the structural modification of the aglycon in chromogenic substrates. Another

explanation could be the production by the hit clones of cell wall anchors or glycoprotein adhesins (as further explained in this paper regarding 9D11 sequence annotation), which may compete for binding with mucin-specific lectins.

## Identification of Mucin Glycan Utilization Loci

The metagenomic DNA from the 13 clones was sequenced using a high sequencing depth (100X) to ensure reliable sequence assembly. For each clone, one contig of between 27,897 and 50,904 bp in size was obtained. In total, 794,676 bp (containing 316 ORFs) were analyzed. Of these 316 ORFs, 56 were annotated as CAZyme encoding genes.

Forty-nine of these predicted CAZymes contained a GH module (**Supplementary Table S1**), and a further four were esterases. In most of the contig sequences, the CAZyme-encoding genes were organized in operon-like multigenic systems similar to the polysaccharide utilization loci (PULs) described for cultivated species (Terrapon et al., 2018) and other metagenomic DNA fragments (Tasse et al., 2010; **Supplementary Table S1**). The metagenomic PULs retrieved in the present study code for the batteries of CAZymes, carbohydrate sensing/binding proteins and transporters required to achieve the complete breakdown and uptake of mucin glycans. The structural modification of mucin glycans results from the production of several glycosidase activities that could be inferred from the GH families highlighted in their sequences (**Table 2** and **Supplementary Table S1**), and which are all required for the degradation of complex

mucin glycans. Most of the GH sequences presented the signal peptide sequences, indicating their membrane or extracellular location in the bacteria they originate from, necessary to initiate the breakdown of complex glycan structures into simpler oligosaccharides, which are very likely internalized thanks to the various carbohydrate transporters we identified (annotated as ABC transport family, MFS, SusD/SusC systems or outer membrane efflux protein). As with canonical *Bacteroidetes* PULs, we also identified other genes coding for proteins involved in the regulation of gene expression in these metagenomic loci (AraC family transcriptional regulator, chemotaxis protein CheY/response regulator receiver domain protein). Finally, as described in greater detail later in this section, we also identified in these loci several genes coding for putative enzymes involved in central carbohydrate metabolism, especially sialic acid metabolism.

Partial or complete sequence redundancy was observed in clones 14N11, 20L12, 47G11, and 39E18, in clones 3I21, 40B3 41E6, and 39O22, in clones 12O6 and 47C24, and in clones 13P9 and 33D18 (**Figures 2A–D**, respectively). These 12 sequences finally correspond to four loci, all assigned to the *Bacteroides* genus using PhyloPythiaS which is in agreement with the synteny detailed hereafter.

### Locus A Assigned to *Bacteroides vulgatus* or *Bacteroides massiliensis*

The three partially redundant sequences of the sialidase-,  $\beta$ -D-NAcetylgalactosaminidase- and/or  $\beta$ -D-NAcetylglucosaminidase-producing clones 14N11, 20L12, and 39E18 presented high sequence identity (99–100%) onto their coverage areas (**Figure 2A**). The 39E18 contig is included in that of 20L12. The 14N11 and 20L12 sequences can be combined to form a longer, single contig containing eight GH-encoding genes assigned to the GH2 (2 sequences), GH20 (3 sequences), GH27, GH33, and GH92 families. These families all host already characterized members which are able to depolymerize human N- and O-glycans;  $\beta$ -galactosidases,  $\beta$ -N-acetylglucosaminidases,  $\alpha$ -N-acetylgalactosaminidases, sialidases and  $\alpha$ -mannosidases, respectively. Synteny was detected between contigs 14N11, 20L12, and 39E18 and a locus from *Bacteroides vulgatus* ATCC 8482 corresponding to the predicted PUL 67. The synteny with PUL 67 is, however, disrupted by the insertion of four genes in the metagenomic DNA (ORFs 13–16 in clone 20L12), including the GH27 encoding gene and two genes involved in nutrient binding and transport. This insertion confers an ecological advantage to the uncultured strain containing this metagenomic locus over *B. vulgatus* for O-GalNAc glycan foraging, especially the core 3 structure from human colonic mucins (Brockhausen et al., 2009). Sialidase activity was observed for contig 47G11 on the chromogenic substrate and mucus. This contig contained a similar gene series to 14N11 and 20L12, albeit with two additional CAZyme-encoding genes (GH29 and 97), and shared, respectively 82 and 90% sequence identity with contigs 14N11 and 20L12. The percentage of its CAZyme sequences' identity with the homologous sequences in contigs 14N11 and 20L12 (79–90%) indicates that these metagenomic fragments

originate from different strains or species. Contig 47G11 indeed presents an almost perfect synteny with the predicted PUL 4 from *B. massiliensis* B84634 (100% Cov, 99% ID), but with two additional genes in the metagenomic sequence.

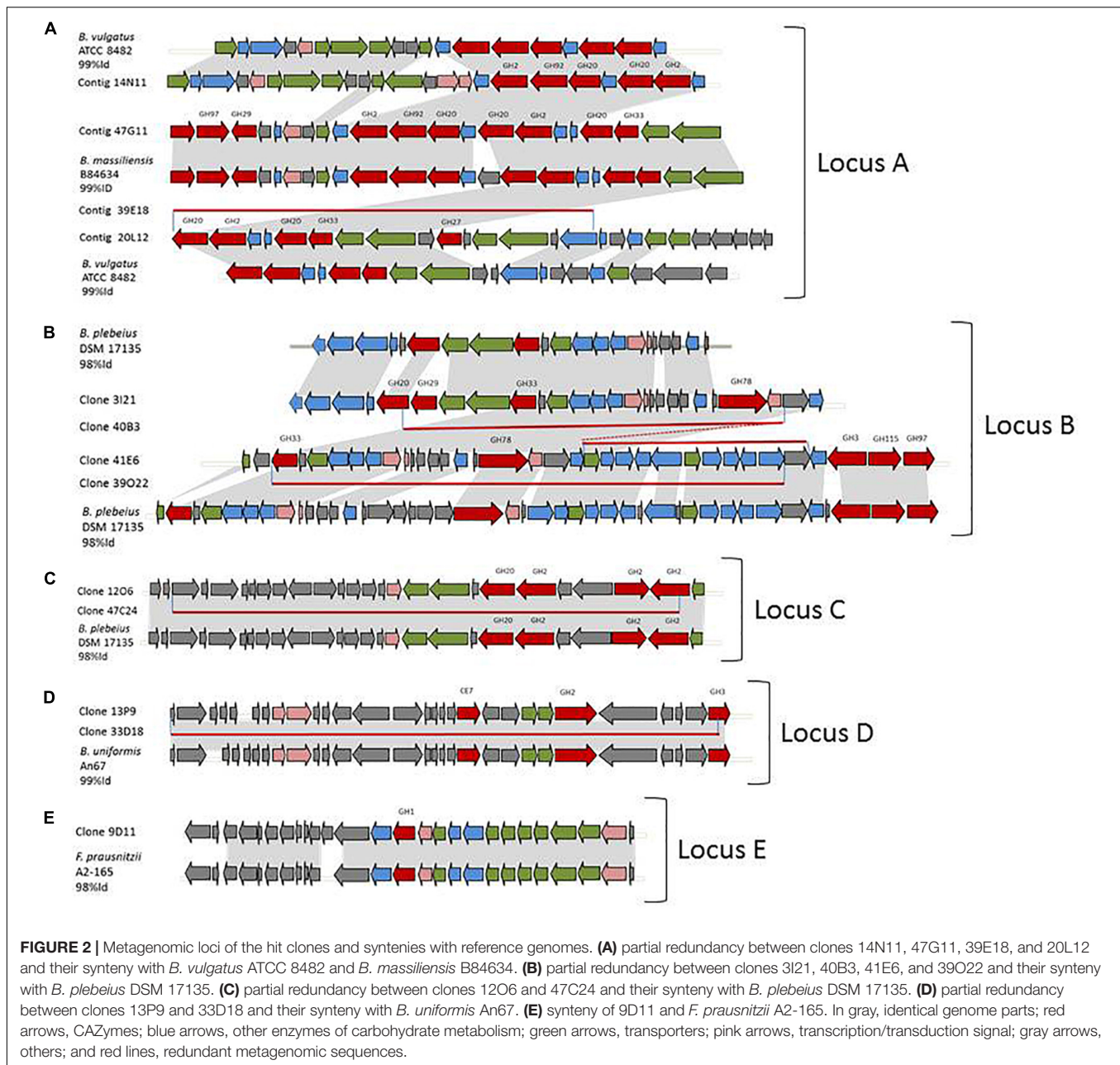
The *B. vulgatus* strain ATCC 8482 is one of the most prevalent strains in the human gut *Bacteroides* group. It is thought to be responsible for the development of inflammation and exacerbated immune response in the pathogenesis of IBD (Shiba et al., 2003). However, in our study, we found the genes of contigs 20L12 and 14N11 to be highly prevalent and abundant in the human gut metagenomic gene catalog, regardless of the geographic origin or medical status of the subjects (**Figure 3** and **Supplementary Table S1**). The locus identified in this study, which is not a biomarker of IBD, is thus unlikely to be involved in IBD. Besides, the *B. vulgatus* strain ATCC 8482, which is known to produce a whole panel of glycolytic activities required to degrade mucus glycoproteins (Ruseler-van Embden et al., 1989), has previously been found to degrade pig gastric mucin, but not human mucin glycans (Hoskins et al., 1992; Png et al., 2010). These data could thus be revisited by more specifically assessing the growth of this strain on specific structures of sialylated mucin glycans. As *B. vulgatus* strain ATCC8482, *B. massiliensis* B84634 was also reported to be able to grow on pig gastric mucin, resulting in the activation of six of its 33 PULs, including PUL 4 (Pudlo et al., 2015). The prevalence of the cluster's genes is lower than those of corresponding genes in the *B. vulgatus* strain, and their abundance is lower in European subjects (**Figure 3** and **Supplementary Table S1**).

Finally, in the contigs of these sialidase-producing clones, we also identified genes encoding putative sialic acid-specific 9-O-acetylsterase, N-acetylglucosamine-6-phosphate deacetylase, glucosamine-6-phosphate deaminase and sialic acid-induced mutarotase, constituting the denominated Nan system for N-acylneuraminate (Vimr et al., 2004; Egan et al., 2014). Marcobal et al. (2011) stressed that a certain number of bacterial species do not possess all the enzymes for the complete catabolism of sialic acid: *B. tetaiotaomicron*, for example, has sialidases but not the Nan system to metabolize sialic acid. Conversely, other species such as *Clostridium difficile* do not have sialidases but possess the Nan system to metabolize sialic acid released by other bacteria. Here, the uncultured bacteria from which contigs 14N11, 20L12, and 39E18 originated evolved to cluster all the sialidase-encoding genes and the Nan operon on the same locus, as well as other CAZyme-encoding genes likely to release other monosaccharides belonging to the sialylated mucin glycans. This confers an important ecological advantage to colonize the mucosal layer.

### Locus B Assigned to *Bacteroides plebeius*

The four partially redundant sequences 3I21, 40B3, 39O22, and 41E6 have high sequence similarity (99–100%) (**Figure 2B**). The 3I21 and 41E6 sequences can be combined to form a longer, single contig containing seven GH-encoding genes, including the GH20, GH29, GH33, GH3 families, which are known to contain members able to depolymerize O-glycans

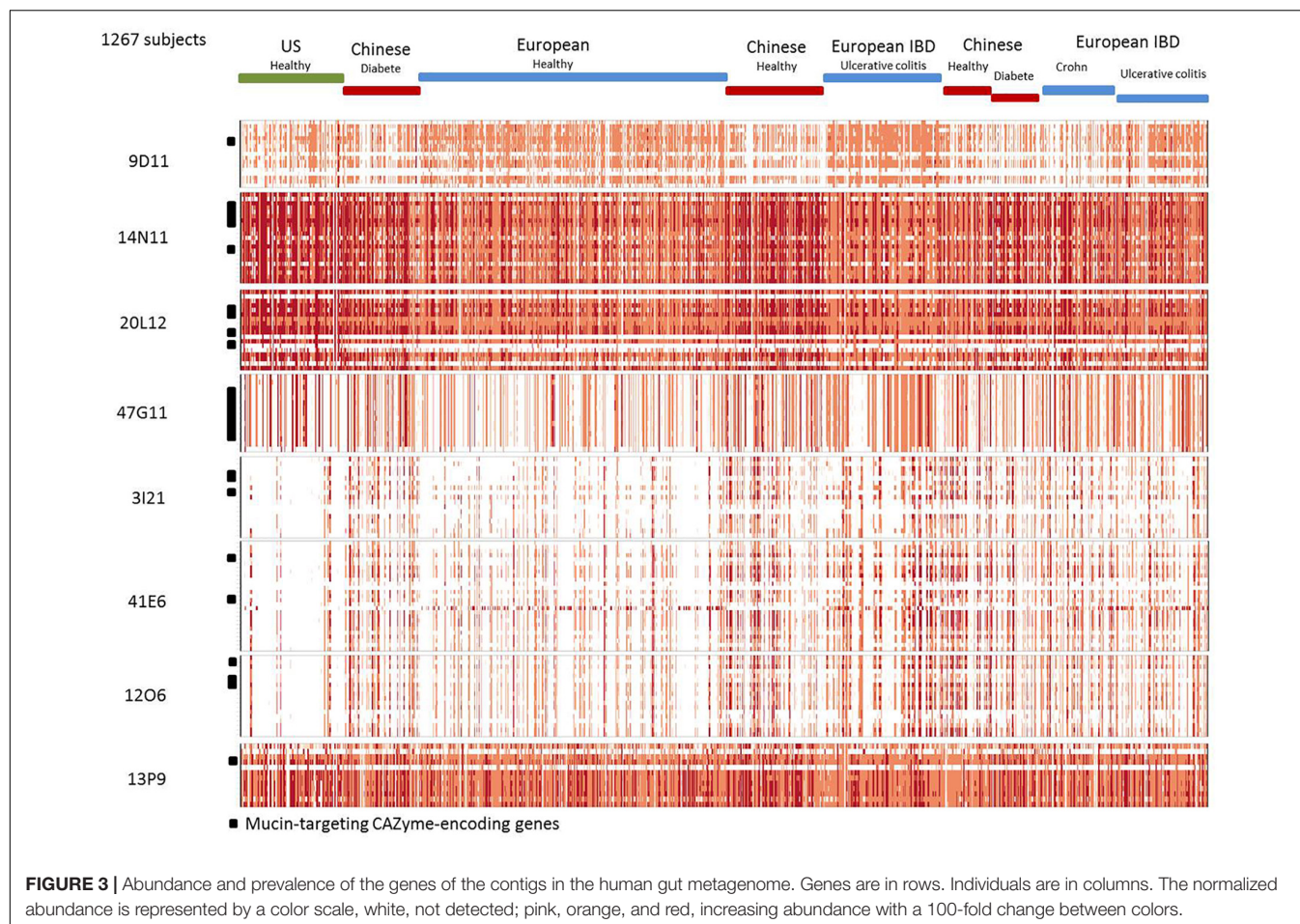




**FIGURE 2 |** Metagenomic loci of the hit clones and synteny with reference genomes. **(A)** partial redundancy between clones 14N11, 47G11, 39E18, and 20L12 and their synteny with *B. vulgatus* ATCC 8482 and *B. massiliensis* B84634. **(B)** partial redundancy between clones 3I21, 40B3, 41E6, and 39O22 and their synteny with *B. plebeius* DSM 17135. **(C)** partial redundancy between clones 12O6 and 47C24 and their synteny with *B. plebeius* DSM 17135. **(D)** partial redundancy between clones 13P9 and 33D18 and their synteny with *B. uniformis* An67. **(E)** synteny of 9D11 and *F. prausnitzii* A2-165. In gray, identical genome parts; red arrows, CAZymes; blue arrows, other enzymes of carbohydrate metabolism; green arrows, transporters; pink arrows, transcription/transduction signal; gray arrows, others; and red lines, redundant metagenomic sequences.

( $\beta$ -1,6-N-acetylglucosaminidases,  $\alpha$ -L-fucosidases, sialidases). The remaining GH78-, GH115-, and GH97-encoding genes belong to another gene cluster best suited to the degradation of plant glycans, which highlights the probable metabolic flexibility of the uncultured bacterium from which this metagenomic DNA fragment originated. The functional profile of the clones matches their content in GH families (Table 2 and Figure 1). Additional genes involved in carbohydrate binding and transport (annotated as substrate-binding protein, SusC/SusD, and major facilitator family transporter) and gene expression regulation were also present in these metagenomic sequences. These contigs present a nearly perfect synteny with the predicted PUL 9 from the *B. plebeius* DSM 17135 strain (93% Cov, 98%

Id for 3I21 and 95% Cov, 98% Id for 41E6), although this was disrupted by the insertion of two genes coding for a GH29 putative  $\alpha$ -L-fucosidase (ORF6 in contig 3I21) and one putative transporter (ORF18 in contig 3I21), adding a complementary function to the bacterium's glycan degradation pathway. In this mucin-targeting metagenomic PUL, GH-encoding genes were also surrounded by genes conferring complementary functions for N-acetylhexosamine catabolism and sialylation processes – methyl transferase, N-acetylneuraminase cytidyltransferase, N-acetylneuraminic mutarotase, sialic acid-specific 9-O-acetyltransferase, N-acetylneuraminase lyase and N-acetylglucosamine 2-epimerase – whose roles in carbohydrate metabolism have previously been described for symbiotic



and pathogenic bacteria (Vimr and Lichtensteiger, 2002; Ringenberg et al., 2003; Bravo et al., 2004; Vimr et al., 2004; Mizanur and Pohl, 2008). This gene cluster thus combines genes dedicated to the depolymerization of sialylated mucin glycans with genes involved in the synthesis, activation and transfer of sialic acid onto bacterial cell surfaces. Such a process of carbohydrate harvesting and reuse for bacterial glycan synthesis would permit bacteria to mimic vertebrate cell surfaces and evade the host's immune system. In mammals, sialic acids are widespread molecules usually at the terminal position of oligosaccharide chains of cell-surface or serum glycoconjugates. Their functions in regulation of host innate defense mechanisms make it a key substance for microorganisms to survive in the environment (Vimr et al., 2004). Although they were all prevalent in the sample, these metagenomic genes are not evenly distributed among the population (Figure 3). Except for one gene (ORF1, contig 41E6), a mobile element appearing in the heat map as a biomarker of European subjects, the entire locus (including the additional GH29 compared to *B. plebeius* DSM 17135 PUL9) is significantly more abundant in the European IBD cohort than the healthy cohort. At the same time, it is abundant and prevalent in the Chinese subjects, including the healthy ones, and almost absent in the United States subjects.

### Locus C Assigned to *Bacteroides plebeius*

Contig 47C24 is included in contig 12O6, with which it shares 99% sequence identity. Contig 12O6 has 98% identity with a *B. plebeius* DSM 17135 locus, which includes a part of the predicted PUL 4 (Figure 2C). This part of PUL 4 contains one GH20- and three GH2-encoding genes, with the former gene explaining the  $\beta$ -D-Nacetylgalactosaminidase and  $\beta$ -D-Nacetylglucosaminidase activities of clones 47C24 and 12O6. Surrounding the CAZymes, we also found three genes likely involved in the binding and transport of nutrients. Finally, not surprisingly, the prevalence and abundance pattern of this locus was similar to that of contigs 3I21, 40B3, 39O22, and 41E6, which also showed synteny with other *B. plebeius* DSM 17135 loci.

### Locus D Assigned to *Bacteroides uniformis*

The last *Bacteroides* contig, namely contig 13P9, which includes sequence 33D18 (Figure 2D), shared 99% identity with a genomic locus from *B. uniformis* An67, a strain which is not included in the PUL database. This metagenomic PUL contains a GH2-, a GH3-, and a CE7-encoding gene. The GH2 family contains  $\beta$ -D-mannosidases, explaining activity on X- $\beta$ -D-Man (Table 2 and



**Figure 1).** Due to its position at the end of the contig, the GH3-encoding gene is truncated and probably not functional here, but some of the characterized proteins of the family are known to act on  $\beta$ -N-acetylhexosamines, of which the acetyl groups could be hydrolysed by the CE7 enzyme. This metagenomic locus is probably involved in the catabolism of human N-glycans by uncultured bacteria, and/or *B. uniformis* strains, which have a high potential to utilize both dietary (Tasse et al., 2010; Patrascu et al., 2017) and endogenous glycans (Benítez-Páez et al., 2017). This considerable metabolic flexibility of *B. uniformis*, which gives it a significant advantage when it comes to easily colonizing the human gut ecosystem, could explain why the locus identified in this study is so prevalent and abundant in the microbiome, regardless of the geographic origin or medical status of the individual (Supplementary Table S1 and Figure 3).

### Locus E Assigned to *Faecalibacterium prausnitzii*

The last metagenomic locus identified here, that of the  $\beta$ -D-mannosidase and  $\beta$ -N-acetylglucosaminidase clone 9D11, was assigned to *Faecalibacterium* using PhylophthiaS. It was indeed found to share 98% identity with a locus of *Faecalibacterium prausnitzii* A2-165 F (Figure 2E). It contains only one CAZyme-encoding gene (GH1), which was probably responsible for the activity detected on X- $\beta$ -Man. Conversely, neither activity on X- $\beta$ -GlcNAc nor modification of WGA and SNA lectin binding to mucus could be explained by the presence of this GH1 enzyme, indicating that one or several of the numerous putative proteins encoded in this locus (some presenting distant homologies with adhesins and agglutinins) may perform these functions, and/or that the encoded GH1 is a highly promiscuous enzyme, which will have to be characterized. Additionally, five genes coding for a putative peptidic transport system (oppABCDF operon) were clustered in this locus. The opp transport system is involved in peptide uptake for nutrition, sensing environmental changes and recycling peptides from, for instance, muropeptides; that is, peptidoglycans from bacterial cell walls consisting of alternating residues of  $\beta$ -N-acetylglucosamine (NAG) and N-acetylmuramic acid (NAM) linked to a peptide chain (Monnet, 2003). In this pathway, peptides internalized by the oppABCDF system are hydrolysed into amino acids by a set of peptidases, a role that may have been performed here by the putative CocE/NonD hydrolase belonging to the aminopeptidase family identified in the 9D11 locus. With respect to the functioning of *F. prausnitzii* in the human gut microbiota, most strains grew on N-acetylglucosamine, highlighting their probable ability to utilize host-derived substrates (Lopez-Siles et al., 2012). This is consistent with the CAZy gene content of the *F. prausnitzii* genomes. We indeed found several *F. prausnitzii* sequences in the CAZy database that had been assigned to families as GH2- and GH3-containing members active on  $\beta$ -D-N-acetylglucosamine,  $\beta$ -D-N-acetylhexosamine and  $\beta$ -D-glucosamine linkages, although none of these *F. prausnitzii* enzymes has been biochemically characterized to date. In any case, *F. prausnitzii* is far from being as well equipped with glycan-degrading enzymes as the *Bacteroides* species described

above. Nevertheless, our results demonstrate that this species has the potential to interact with mucus. This could explain the above mentioned modulation of the effect of *B. thetaiotaomicron* on the intestinal mucus barrier, resulting in a modification in goblet cell differentiation and mucin glycosylation (Wrzosek et al., 2013). Moreover, *F. prausnitzii* belongs to the most abundant Firmicutes species in the human gut, which is consistent with the high abundance and prevalence of the locus we identified in this study (Figure 3). The gene abundance heat map indicates a lower prevalence of 9D11 genes in Chinese subjects, regardless of their medical status, than in European ones. What is more, compared to healthy subjects, these genes are more abundant in the microbiome of patients suffering from ulcerative colitis, but less abundant in patients with Crohn's disease (CD). This is perfectly consistent with the low abundance of *F. prausnitzii* in the microbiome of CD patients, particularly those with ileal involvement (Sokol et al., 2009), which could be connected with the ability of this species to interact with mucus through its glycan and proteic fractions, as highlighted here with the 9D11 locus.

### CONCLUSION

To conclude, our work constitutes the first activity-based metagenomic study targeting the glycan-mediated relationships between the human gut microbiota and the host. Until now, the identification of bacterial species and metabolic pathways involved in mucus degradation was restricted to cultured bacteria, tested for growth on complex mucin glycan structures of porcine origin. When combined to genomic and transcriptomic analyses, these studies allowed to identify mucin glycan utilization loci for the few dozen of targeted strains. Nevertheless, to date only a very small number of these pathways have been biochemically proven to degrade human mucin structures.

In the present study, we designed a two-step functional screening strategy, which allowed us, for the first time, both to identify mucus degrading pathways from uncultured species of the microbiota, and to prove their ability to degrade specific motives of human colonic mucus. The mucin glycan utilization loci we discovered are highly prevalent and abundant in the microbiome, and present marked syntenies with loci from several prominent commensal strains from the Bacteroidetes and Firmicutes phyla, most of them being known to forage on dietary glycans, but which had never been described to feed on mucin glycans. Our data highlight the metabolic flexibility of these abundant commensals, which would have the ability to easily redirect their metabolism from dietary polysaccharides to host-derived glycans, depending on the availability of resources. In addition, we highlighted the overabundance of certain of these (meta) genomic loci in the microbiome of patients suffering from inflammatory diseases. This particular trait, together with the proof of their involvement in the modification of human mucus, makes them new targets to study the host response to the degradation of the mucus firewall, and to elaborate new strategies to restore gut homeostasis, for example by using specific functional foods to divert over-abundant mucus foragers toward dietary glycans.

## ETHICS STATEMENT

Bacterial metagenomic DNA was obtained from the distal ileum sample of a 51-year-old male patient undergoing colonoscopy and surgery for a suspected cancer of the lower colon after he had been subjected to a cleansing preparation. The patient did not receive any antibiotics or other drugs in the 6 months before sampling. The sampling protocol was approved by the local ethics committee: the Comité de Protection des Personnes Sud Est V (Ref.: 07-CHUG-21; B. Habozit, J. L. Crakowski, J. Juge, J. Grunwald, E. Svhan, and E. Fontaine on the committee) and informed written consent was obtained from the subject before sampling. The human colon mucus was isolated from patients (Ajandouz et al., 2016). Briefly, human intestinal colonic resections (5–10 cm long) were obtained from patients undergoing surgery for colonic adenocarcinoma at the Hospital Nord of Marseille's general and gastrointestinal surgery unit as part of a collaborative "clinical transfer" project. The sampling procedures were approved by the French Ethics Committee (CODECOH no. DC-2011-1319).

## AUTHOR CONTRIBUTIONS

EL conceived the study, designed the work, acquired, analyzed and interpreted the data, and drafted the work. JP and MM

conceived the study, designed the work, acquired and analyzed the data, and drafted the final revision of the manuscript. NB and JE acquired, analyzed and interpreted the data, and drafted the final revision of the manuscript. AT, EB, ED, and BH analyzed and interpreted the data, and drafted the final revision of the manuscript. ML, SB, EDP, and PR conceived the study and designed the work, and drafted the final revision of the manuscript. GP-V conceived the study, designed the work, analyzed and interpreted the data, and drafted the final revision of the manuscript.

## FUNDING

This research was funded by the French Ministry of Education and Research (Ministère de l'Enseignement Supérieur et de la Recherche), the INRA's M2E Metaprogramme (the Metascreen project), and the European Union's Horizon 2020 Framework Program (LEITBIO-2015-685474, Metafluidics).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2019.01286/full#supplementary-material>

## REFERENCES

- Ajandouz, E. H., Berdah, S., Moutardier, V., Bege, T., Birnbaum, D. J., Perrier, J., et al. (2016). Hydrolytic fate of 3/15-acetyldeoxynivalenol in humans: specific deacetylation by the small intestine and liver revealed using in Vitro and ex Vivo approaches. *Toxins* 8:232. doi: 10.3390/toxins8080232
- Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. doi: 10.1186/1471-2164-9-75
- Bäckhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A., and Gordon, J. I. (2005). Host-bacterial mutualism in the human intestine. *Science* 307, 1915–1920. doi: 10.1126/science.1104816
- Benítez-Páez, A., Gómez del Pulgar, E. M., and Sanz, Y. (2017). The glycolytic versatility of *Bacteroides uniformis* CECT 7771 and its genome response to oligo and polysaccharides. *Front. Cell Infect. Microbiol.* 7:383. doi: 10.3389/fcimb.2017.00383
- Bravo, I. G., García-Vallvé, S., Romeu, A., and Reglero, Á. (2004). Prokaryotic origin of cytidyltransferases and  $\alpha$ -ketoacid synthases. *Trends Microbiol.* 12, 120–128. doi: 10.1016/j.tim.2004.01.004
- Brockhausen, I., Schachter, H., and Stanley, P. (2009). "O-GalNAc Glycans," in *Essentials of Glycobiology*, eds A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, et al. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).
- Cecchini, D. A., Laville, E., Laguerre, S., Robe, P., Leclerc, M., Doré, J., et al. (2013). Functional metagenomics reveals novel pathways of prebiotic breakdown by human gut bacteria. *PLoS One* 8:e72766. doi: 10.1371/journal.pone.0072766
- Courtois, S., Cappellano, C. M., Ball, M., Francou, F.-X., Normand, P., Helync, G., et al. (2003). Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Appl. Environ. Microbiol.* 69, 49–55. doi: 10.1128/aem.69.1.49-55.2003
- Derrien, M., Vaughan, E. E., Plugge, C. M., and de Vos, W. M. (2004). *Akkermansia muciniphila* gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *Int. J. Syst. Evol. Microbiol.* 54, 1469–1476. doi: 10.1099/ijs.0.02873-0
- Desai, M. S., Seekatz, A. M., Koropatkin, N. M., Kamada, N., Hickey, C. A., Wolter, M., et al. (2016). A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. *Cell* 167, 1339–1353.e21. doi: 10.1016/j.cell.2016.10.043
- Egan, M., O'Connell Motherway, M., Ventura, M., and van Sinderen, D. (2014). Metabolism of sialic acid by *Bifidobacterium breve* UCC2003. *Appl. Environ. Microbiol.* 80, 4414–4426. doi: 10.1128/AEM.01114-14
- Hoskins, L. C., Boulding, E. T., Gerken, T. A., Harouny, V. R., and Kriaris, M. S. (1992). Mucin glycoprotein degradation by mucin oligosaccharide-degrading strains of human faecal bacteria. Characterisation of saccharide cleavage products and their potential role in nutritional support of larger faecal bacterial populations. *Microb. Ecol. Health Dis.* 5, 193–207. doi: 10.3109/08910609209141586
- Jakobsson, H. E., Rodríguez-Piñero, A. M., Schütte, A., Ermund, A., Boysen, P., Bemark, M., et al. (2015). The composition of the gut microbiota shapes the colon mucus barrier. *EMBO Rep.* 16, 164–177. doi: 10.15252/embr.201439263
- El Kaoutari, A., Armougom, F., Gordon, J. I., Raoult, D., and Henrissat, B. (2013). The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat. Rev. Microbiol.* 11, 497–504. doi: 10.1038/nrmicro3050
- Katayama, T., Fujita, K., and Yamamoto, K. (2005). Novel bifidobacterial glycosidases acting on sugar chains of mucin glycoproteins. *J. Biosci. Bioeng.* 99, 457–465. doi: 10.1263/jbb.99.457
- Ladevèze, S., Cioci, G., Roblin, P., Mourey, L., Tranier, S., and Potocki-Véronèse, G. (2015). Structural bases for N-glycan processing by mannoside phosphorylase. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 71, 1335–1346. doi: 10.1107/S1399004715006604
- Ladevèze, S., Tarquis, L., Cecchini, D. A., Bercovici, J., Andre, I., Topham, C. M., et al. (2013). Role of glycoside phosphorylases in mannose foraging by human gut bacteria. *J. Biol. Chem.* 288, 32370–32383. doi: 10.1074/jbc.M113.483628
- Li, J., Jia, H., Cai, X., Zhong, H., Feng, Q., Sunagawa, S., et al. (2014). An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotech.* 32, 834–841. doi: 10.1038/nbt.2942
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, D490–D495. doi: 10.1093/nar/gkt1178
- Lopez-Siles, M., Khan, T. M., Duncan, S. H., Harmsen, H. J. M., Garcia-Gil, L. J., and Flint, H. J. (2012). Cultured representatives of two major phylogroups of



- human colonic faecalibacterium prausnitzii can utilize pectin, uronic acids, and host-derived substrates for growth. *Appl. Environ. Microbiol.* 78, 420–428. doi: 10.1128/AEM.06858-11
- Marcobal, A., Barboza, M., Sonnenburg, E. D., Pudlo, N., Martens, E. C., Desai, P., et al. (2011). *Bacteroides* in the infant gut consume milk oligosaccharides via mucus-utilization pathways. *Cell Host Microbe* 10, 507–514. doi: 10.1016/j.chom.2011.10.007
- Marcobal, A., Southwick, A. M., Earle, K. A., and Sonnenburg, J. L. (2013). A refined palate: bacterial consumption of host glycans in the gut. *Glycobiology* 23, 1038–1046. doi: 10.1093/glycob/cwt040
- Martens, E. C., Chiang, H. C., and Gordon, J. I. (2008). Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* 4, 447–457. doi: 10.1016/j.chom.2008.09.007
- Martens, E. C., Neumann, M., and Desai, M. S. (2018). Interactions of commensal and pathogenic microorganisms with the intestinal mucosal barrier. *Nat. Rev. Microbiol.* 16, 457–470. doi: 10.1038/s41579-018-0036-x
- Mizanur, R. M., and Pohl, N. L. (2008). Bacterial CMP-sialic acid synthetases: production, properties, and applications. *Appl. Microbiol. Biotechnol.* 80, 757–765. doi: 10.1007/s00253-008-1643-7
- Monnet, V. (2003). Bacterial oligopeptide-binding proteins. *Cell. Mol. Life Sci.* 60, 2100–2114. doi: 10.1007/s00018-003-3054-3
- Nielsen, H. (2017). “Predicting Secretory Proteins with SignalP” in *Protein Function Prediction: Methods and Protocols*, ed. D. Kihara (New York, NY: Springer), 59–73. doi: 10.1007/978-1-4939-7015-5\_6
- Owen, C. D., Tailford, L. E., Monaco, S., Šuligoj, T., Vaux, L., Lallement, R., et al. (2017). Unravelling the specificity and mechanism of sialic acid recognition by the gut symbiont *Ruminococcus gnavus*. *Nat. Commun.* 8:2196. doi: 10.1038/s41467-017-02109-8
- Patrascu, O., Béguet-Crespel, F., Marinelli, L., Le Chatelier, E., Abraham, A.-L., Leclerc, M., et al. (2017). A fibrolytic potential in the human ileum mucosal microbiota revealed by functional metagenomic. *Sci. Rep.* 7:40248. doi: 10.1038/srep40248
- Pickard, J. M., Maurice, C. F., Kinnebrew, M. A., Abt, M. C., Schenten, D., Golovkina, T. V., et al. (2014). Rapid fucosylation of intestinal epithelium sustains host–commensal symbiosis in sickness. *Nature* 514, 638–641. doi: 10.1038/nature13823
- Png, C. W., Lindén, S. K., Gilshenan, K. S., Zoetendal, E. G., McSweeney, C. S., Sly, L. I., et al. (2010). Mucolytic bacteria with increased prevalence in IBD mucosa augment *In Vitro* utilization of mucin by other bacteria. *Am. J. Gastroenterol.* 105, 2420–2428. doi: 10.1038/ajg.2010.281
- Podolsky, D. K. (1985). Oligosaccharide structures of human colonic mucin. *J. Biol. Chem.* 260, 8262–8271.
- Pudlo, N. A., Urs, K., Kumar, S. S., German, J. B., Mills, D. A., and Martens, E. C. (2015). Symbiotic human gut bacteria with variable metabolic priorities for host mucosal glycans. *mBio* 6:e1282-15. doi: 10.1128/mBio.01282-15
- Ringenberg, M. A., Steenbergen, S. M., and Vimr, E. R. (2003). The first committed step in the biosynthesis of sialic acid by *Escherichia coli* K1 does not involve a phosphorylated N-acetylmannosamine intermediate. *Mol. Microbiol.* 50, 961–975. doi: 10.1046/j.1365-2958.2003.03741.x
- Robbe, C., Capon, C., Coddeville, B., and Michalski, J.-C. (2004). Structural diversity and specific distribution of O-glycans in normal human mucins along the intestinal tract. *Biochem. J.* 384, 307–316. doi: 10.1042/bj20040605
- Ruseler-van Embden, J. G., van der Helm, R., and van Lieshout, L. M. (1989). Degradation of intestinal glycoproteins by *Bacteroides vulgatus*. *FEMS Microbiol. Lett.* 58, 37–41. doi: 10.1016/0378-1097(89)90338-8
- Shiba, T., Aiba, Y., Ishikawa, H., Ushiyama, A., Takagi, A., Mine, T., et al. (2003). The suppressive effect of bifidobacteria on *Bacteroides vulgatus*, a putative pathogenic microbe in inflammatory bowel disease. *Microbiol. Immunol.* 47, 371–378. doi: 10.1111/j.1348-0421.2003.tb03368.x
- Sokol, H., Seksik, P., Furet, J. P., Firmesse, O., Nion-Larmurier, I., Beaugerie, L., et al. (2009). Low counts of faecalibacterium prausnitzii in colitis microbiota. *Inflamm. Bowel Dis.* 15, 1183–1189. doi: 10.1002/ibd.20903
- Sonnenburg, J. L., Xu, J., Leip, D. D., Chen, C.-H., Westover, B. P., Weatherford, J., et al. (2005). Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science* 307, 1955–1959. doi: 10.1126/science.1109051
- Tailford, L. E., Crost, E. H., Kavanaugh, D., and Juge, N. (2015). Mucin glycan foraging in the human gut microbiome. *Front. Genet.* 6:81. doi: 10.3389/fgene.2015.00081
- Tasse, L., Bercovici, J., Pizzut-Serin, S., Robe, P., Tap, J., Klopp, C., et al. (2010). Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. *Genome Res.* 20, 1605–1612. doi: 10.1101/gr.108332.110
- Terrapon, N., Lombard, V., Drula, É., Lapébie, P., Al-Masaudi, S., Gilbert, H. J., et al. (2018). PULDB: the expanded database of polysaccharide utilization loci. *Nucleic Acids Res.* 46, D677–D683. doi: 10.1093/nar/gkx1022
- Vimr, E., and Lichtensteiger, C. (2002). To sialylate, or not to sialylate: that is the question. *Trends Microbiol.* 10, 254–257. doi: 10.1016/s0966-842x(02)02361-2
- Vimr, E. R., Kalivoda, K. A., Deszo, E. L., and Steenbergen, S. M. (2004). Diversity of microbial sialic acid metabolism. *Microbiol. Mol. Biol. Rev.* 68, 132–153. doi: 10.1128/mmbr.68.1.132-153.2004
- Wrzosek, L., Miquel, S., Noordine, M.-L., Bouet, S., Chevalier-Curt, M. J., Robert, V., et al. (2013). *Bacteroides thetaiotaomicron* and *Faecalibacterium prausnitzii* influence the production of mucus glycans and the development of goblet cells in the colonic epithelium of a gnotobiotic model rodent. *BMC Biol.* 11:61. doi: 10.1186/1741-7007-11-61

**Conflict of Interest Statement:** PR was employed by company Givaudan SA.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Laville, Perrier, Bejar, Maresca, Esque, Tauzin, Bouhajja, Leclerc, Drula, Henrissat, Berdah, Di Pasquale, Robe and Potocki-Veronese. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.