



Supplementary Materials

Arabinoxylan and Pectin Metabolism in Crohn's Disease Microbiota: An *In Silico* Study

Carlos Sabater ^{1,2,*}, Inés Calvete-Torre ^{1,2}, Lorena Ruiz ^{1,2} and Abelardo Margolles ^{1,2}

¹ Department of Microbiology and Biochemistry of Dairy Products, Instituto de Productos Lácteos de Asturias-Consejo Superior de Investigaciones Científicas (IPLA-CSIC), Paseo Río Linares s/n, 33300 Villaviciosa, Asturias, Spain; ines.calvete@ipla.csic.es (I.C.-T.); lorena.ruiz@ipla.csic.es (L.R.); amargolles@ipla.csic.es (A.M.)

² Functionality and Ecology of Beneficial Microbes (MicroHealth) Group, Instituto de Investigación Sanitaria del Principado de Asturias (ISPA), 33011 Oviedo, Asturias, Spain

* Correspondence: carlos.sabater@csic.es; Tel.: +34-985-892131

Table S1. Number of metagenome-assembled genomes (MAGs, n=198) recovered only from the microbiota of patients with Crohn's disease. MAGs were identified at family, genus, species or strain level (total number of different taxa n=25).

MAGs found only in the microbiota of patients with Crohn's disease	
Taxa	n
<i>Parabacteroides distasonis</i>	62
<i>Roseburia intestinalis</i> L1-82	58
<i>Veillonella</i>	24
<i>[Clostridium] bolteae</i>	11
<i>Muribaculum</i> sp. TLL-A4	6
<i>[Clostridium] saccharolyticum</i> WM1	4
<i>[Clostridium] sphenoides</i> JCM 1415	4
<i>Morganella morganii</i>	4
Bacillaceae	3
<i>Proteus mirabilis</i> BB2000	3
<i>Helicobacter canis</i>	2
<i>Klebsiella variicola</i>	2
<i>Proteus mirabilis</i> HI4320	2
<i>Roseburia hominis</i>	2
<i>[Clostridium] scindens</i> ATCC 35704	1
<i>Actinomyces</i> sp. oral taxon 414	1
<i>Bifidobacterium asteroides</i>	1
<i>Bifidobacterium dentium</i>	1
<i>Flintibacter</i> sp. KGMB00164	1
<i>Fusobacterium necrophorum</i> subsp. <i>funduliforme</i>	1
<i>Massilistercora timonensis</i>	1
<i>Muribaculum</i> sp. H5	1
<i>Olsenella</i> sp. GAM18	1
Rhodobacteraceae	1
<i>Streptococcus lutetiensis</i> 033	1
Total	198

Table S2. Number of metagenome-assembled genomes (MAGs, n=216) recovered only from the microbiota of healthy individuals. MAGs were identified at family, genus, species or strain level (total number of different taxa n=42).

MAGs found only in the microbiota of healthy individuals	
Taxa	n
<i>Parabacteroides</i>	54
<i>Clostridium</i>	34
<i>Roseburia</i>	20
<i>Muribaculum</i>	18
<i>Bifidobacterium adolescentis</i> ATCC 15703	14
<i>Acidaminococcus fermentans</i> DSM 20731	6
<i>Brachyspira murdochii</i> DSM 12563	5
<i>Desulfovibrio piger</i>	5
<i>Faecalitalea cylindroides</i> T2-87	5
<i>Lactobacillus ruminis</i> ATCC 27782	4
<i>Methanobrevibacter smithii</i> ATCC 35061	4
<i>Streptococcus thermophilus</i>	4
<i>Veillonella atypica</i>	4
<i>Clostridioides difficile</i>	3
<i>Lactobacillus mucosae</i> LM1	3
<i>Veillonella parvula</i> HSIVP1	3
Burkholderiaceae	2
<i>Cutibacterium acnes</i>	2
<i>Lactobacillus fermentum</i>	2
<i>Turicibacter</i> sp. H121	2
<i>Absiella</i>	1
Acidaminococcaceae	1
<i>Adlercreutzia equolifaciens</i> DSM 19450	1
<i>Bifidobacterium animalis</i> subsp. <i>animalis</i>	1
<i>Bifidobacterium bifidum</i> BGN4	1
<i>Butyricimonas faecalis</i>	1
<i>Campylobacter concisus</i> 13826	1
Clostridiales Family XIII. Incertae Sedis	1
<i>Fusobacterium ulcerans</i> ATCC 49185	1
<i>Lactobacillus amylovorus</i> DSM 20531	1
<i>Lactobacillus salivarius</i> str. Ren	1
Methanomassiliicoccaceae archaeon DOK	1
Muribaculaceae bacterium DSM 108610	1
<i>Oscillibacter</i> sp. PEA192	1
Oscillospiraceae	1
<i>Paenibacillus</i>	1
Rhodospirillaceae	1
<i>Selenomonas sputigena</i> ATCC 35185	1
<i>Streptococcus salivarius</i> JIM8777	1
<i>Succinivibrio dextrinosolvens</i>	1
<i>Veillonella dispar</i>	1
Victivallales bacterium CCUG 44730	1
Total	216

Table S3. Enzyme domains acting on arabinoxylan (AX) and pectin annotated in metagenome-assembled genomes (MAGs) recovered from the microbiota of patients with Crohn's disease. These MAGs were assigned to taxonomic clades that were also found in the microbiota of healthy individuals.

Domain	Acting on	Potential activities of interest
CBM2	AX	Xylanase
CBM4	AX	Xylanase
CBM35	AX	Xylanase
GH5 2	AX	Endo- β -1,4-xylanase
GH5 4	AX	Endo- β -1,4-xylanase
GH5 21	AX	Endo- β -1,4-xylanase
GH8	AX	Endo- β -1,4-xylanase
GH10	AX	Arabinoxylan-specific endo- β -1,4-xylanase
GH30 8	AX	Glucuronoarabinoxylan-specific endo- β -1,4-xylanase
GH30 1	AX	Endo- β -1,4-xylanase; β -xylosidase
GH120	AX	β -xylosidase
GH115	AX	Xylan α -1,2-glucuronidase
CE1	AX	Acetyl xylan esterase
CE2	AX	Acetyl xylan esterase
CE3	AX	Acetyl xylan esterase
CE4	AX	Acetyl xylan esterase
CE6	AX	Acetyl xylan esterase
CE7	AX	Acetyl xylan esterase
GH3	AX; pectin	Xylan 1,4- β -xylosidase; α -L-arabinofuranosidase
GH43 10	AX; pectin	Xylan 1,4- β -xylosidase; α -L-arabinofuranosidase
GH43 11	AX; pectin	Xylan 1,4- β -xylosidase; α -L-arabinofuranosidase
GH43 12	AX; pectin	Xylan 1,4- β -xylosidase; α -L-arabinofuranosidase
GH43 9	AX; pectin	α -L-arabinofuranosidase
GH43 16	AX; pectin	α -L-arabinofuranosidase
GH43 19	AX; pectin	α -L-arabinofuranosidase
GH43 26	AX; pectin	α -L-arabinofuranosidase
GH43 18	AX; pectin	α -L-arabinofuranosidase
GH43 33	AX; pectin	α -L-arabinofuranosidase
GH43 34	AX; pectin	β -D-galactofuranosidase; α -L-arabinofuranosidase
GH98	AX; pectin	Endo- β -1,4-galactosidase; endo- β -1,4-xylanase
GH1	AX; pectin	β -galactosidase; β -xylosidase
GH2	AX; pectin	β -galactosidase; α -L-arabinofuranosidase; β -xylosidase
GH39	AX; pectin	β -galactosidase; α -L-arabinofuranosidase; β -xylosidase
GH43 1	AX; pectin	β -galactosidase; α -L-arabinofuranosidase; β -xylosidase
CE12	AX; pectin	Acetyl xylan esterase; pectin acetyl esterase

GH42	Pectin	β -galactosidase
CBM32	Pectin	Polygalacturonase
GH78	Pectin	α -L-rhamnosidase
GH105	Pectin	Unsaturated rhamnogalacturonyl hydrolase
CE8	Pectin	Pectin methylesterase
PL11 1	Pectin	Rhamnogalacturonan endo- / exolyase
PL26	Pectin	Rhamnogalacturonan exolyase

Table S4. Enzyme domains acting on arabinoxylan (AX) and pectin annotated in metagenome-assembled genomes (MAGs) recovered from the microbiota of healthy individuals. These MAGs were assigned to taxonomic clades that were also found in the microbiota of patients with Crohn's disease.

Domain	Acting on	Potential activities of interest
CBM4	AX	Xylanase
CBM35	AX	Xylanase
GH5 2	AX	Endo-b-1,4-xylanase
GH5 4	AX	Endo-b-1,4-xylanase
GH5 21	AX	Endo- β -1,4-xylanase
GH5 54	AX	Endo-b-1,4-xylanase
GH8	AX	Endo- β -1,4-xylanase
GH10	AX	Arabinoxylan-specific endo- β -1,4-xylanase
GH30 8	AX	Glucuronoarabinoxylan-specific endo- β -1,4-xylanase
GH120	AX	β -xylosidase
GH115	AX	Xylan α -1,2-glucuronidase
CE1	AX	Acetyl xylan esterase
CE2	AX	Acetyl xylan esterase
CE3	AX	Acetyl xylan esterase
CE4	AX	Acetyl xylan esterase
CE6	AX	Acetyl xylan esterase
CE7	AX	Acetyl xylan esterase
AA10	AX	Xylan oxidase
GH3	AX; pectin	Xylan 1,4- β -xylosidase; α -L-arabinofuranosidase
GH43 10	AX; pectin	Xylan 1,4- β -xylosidase; α -L-arabinofuranosidase
GH43 11	AX; pectin	Xylan 1,4- β -xylosidase; α -L-arabinofuranosidase
GH43 12	AX; pectin	Xylan 1,4- β -xylosidase; α -L-arabinofuranosidase
GH43 27	AX; pectin	Xylan 1,4- β -xylosidase; α -L-arabinofuranosidase
GH43 9	AX; pectin	α -L-arabinofuranosidase
GH43 16	AX; pectin	α -L-arabinofuranosidase
GH43 18	AX; pectin	α -L-arabinofuranosidase
GH43 19	AX; pectin	α -L-arabinofuranosidase
GH43 26	AX; pectin	α -L-arabinofuranosidase
GH43 33	AX; pectin	α -L-arabinofuranosidase
GH43 17	AX; pectin	α -L-arabinofuranosidase; endo- β -1,4-xylanase

GH43 34	AX; pectin	β -D-galactofuranosidase; α -L-arabinofuranosidase
GH98	AX; pectin	Endo- β -1,4-galactosidase; endo- β -1,4-xylanase
GH1	AX; pectin	β -galactosidase; β -xylosidase
GH2	AX; pectin	β -galactosidase; α -L-arabinofuranosidase; β -xylosidase
GH43 1	AX; pectin	β -galactosidase; α -L-arabinofuranosidase; β -xylosidase
CE12	AX; pectin	Acetyl xylan esterase; pectin acylesterase
CBM51	Pectin	Galactosidase
GH42	Pectin	β -galactosidase
CBM32	Pectin	Polygalacturonase
GH28	Pectin	Exopolygalacturonase; rhamnogalacturonase
GH78	Pectin	α -L-rhamnosidase
GH105	Pectin	Unsaturated rhamnogalacturonyl hydrolase
CE8	Pectin	Pectin methylesterase
PL11 1	Pectin	Rhamnogalacturonan endo- / exolyase
PL26	Pectin	Rhamnogalacturonan exolyase
PL9 1	Pectin	Exopolygalacturonate lyase; rhamnogalacturonan endolyase

Table S5. Enzyme domains acting on arabinoxylan (AX), pectin and other carbohydrates (Other) an-notated in metagenome-assembled genomes (MAGs) recovered from the microbiota of patients with Crohn's disease. These enzyme domains were not found in MAGs recovered from metagenome samples from healthy individuals.

Domain	Acting on	Potential activities of interest
CBM54	AX	Xylanase
CBM42	AX; pectin	Arabinofuranosidase
PL4	Pectin	Rhamnogalacturonan endolyase
CBM20	Other	Starch-binding enzyme
CBM46	Other	Cellulase
CBM69	Other	Starch-binding enzyme
PL13	Other	Heparin lyase
PL15 2	Other	Heparin lyase
PL17 1	Other	Oligoalginate lyase

Table S6. Enzyme domains acting on arabinoxylan (AX), pectin and other carbohydrates (Other) annotated in metagenome-assembled genomes (MAGs) recovered from the microbiota of healthy individuals. These enzyme domains were not found in MAGs recovered from metagenome samples from patients with Crohn's disease.

Domain	Acting on	Potential activities of interest
GH11	AX	Endo- β -1,4-xylanase; exo-1,4- β -xylosidase
PL1 6	Pectin	Pectin lyase
PL9 2	Pectin	Pectin lyase
PL2 2	Pectin	Exo-polygalacturonate lyase
PL11 2	Pectin	Rhamnogalacturonan endo- / exolyase
PL4 4	Pectin	Rhamnogalacturonan endolyase
CBM21	Other	Starch-binding enzyme
CBM25	Other	Starch-binding enzyme
CBM63	Other	Cellulase
CBM68	Other	Pullulanase
GH13 33	Other	Trehalose synthase
GH23	Other	Lysozyme type G
GH48	Other	Cellobiohydrolase
GH5 48	Other	Chitosanase
GT87	Other	α -1,2-mannosyltransferase

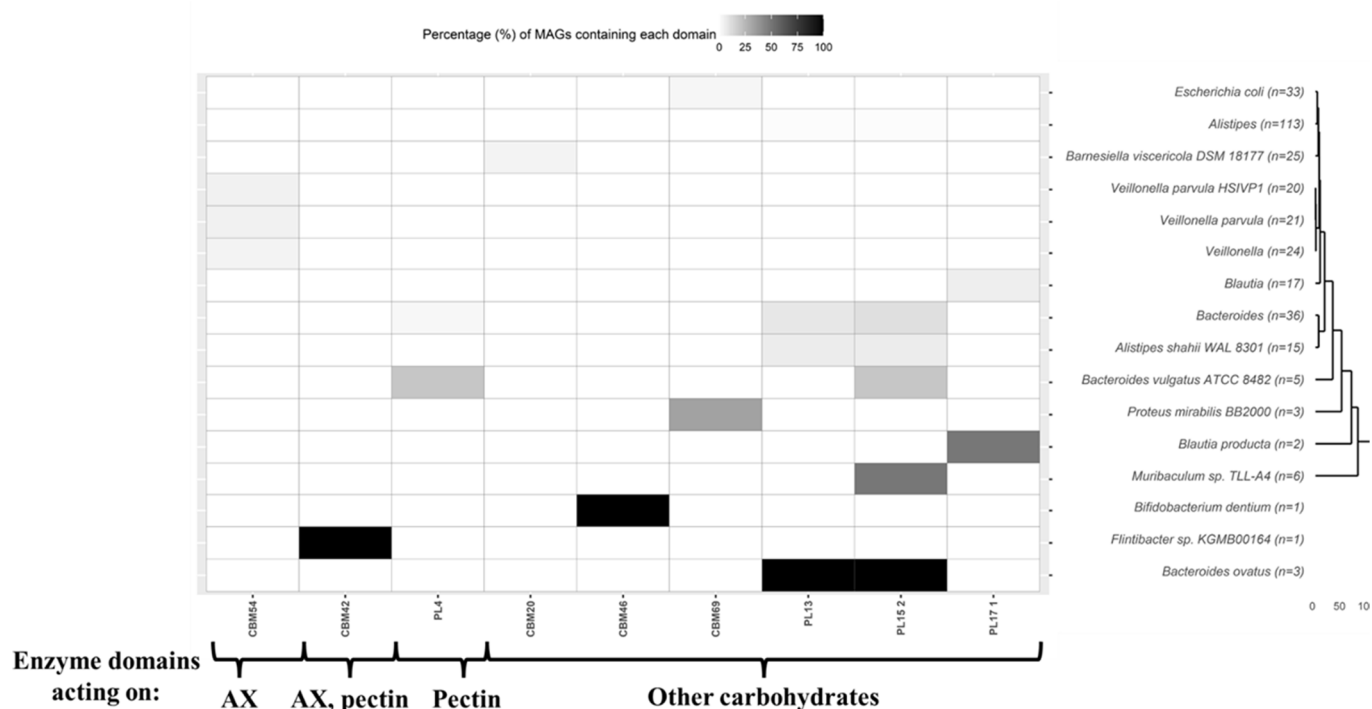


Figure S1. Heatmap showing the presence of different glycosidases (indicated as black cells) in metagenome-assembled genomes (MAGs) recovered only from the microbiota of patients with Crohn's disease (see Table S5). These enzyme domains were not found in MAGs recovered from metagenome samples from healthy individuals. The percentage (%) of MAGs containing each functional domain is shown. Specifically, glycosidases capable of degrading arabinoxylan (AX), pectin and other carbohydrates are illustrated. Glycosidase functional domains showing coverage values higher than 0.95 were annotated. Codes corresponding to the Carbohydrate-Active enZYmes Database (CAZy) family of each enzyme have been assigned.

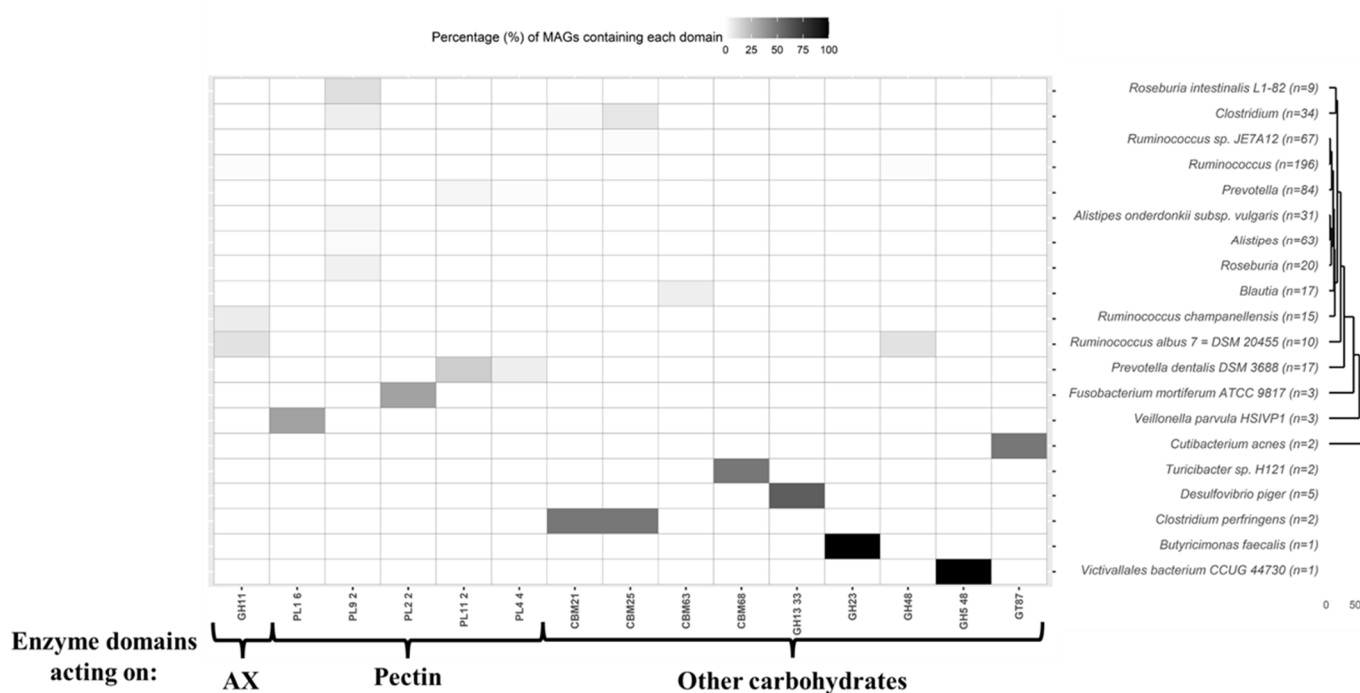


Figure S2. Heatmap showing the presence of different glycosidases (indicated as black cells) in metagenome-assembled genomes (MAGs) recovered only from the microbiota of healthy individuals (see Table S6). These enzyme domains were not found in MAGs recovered from metagenome samples from patients with Crohn's disease. The percentage (%) of MAGs containing each functional domain is shown. Specifically, glycosidases capable of degrading arabinoxylan (AX), pectin and other carbohydrates are illustrated. Glycosidase functional domains showing coverage values higher than 0.95 were annotated. Codes corresponding to the Carbohydrate-Active enZymes Database (CAZy) family of each enzyme have been assigned.

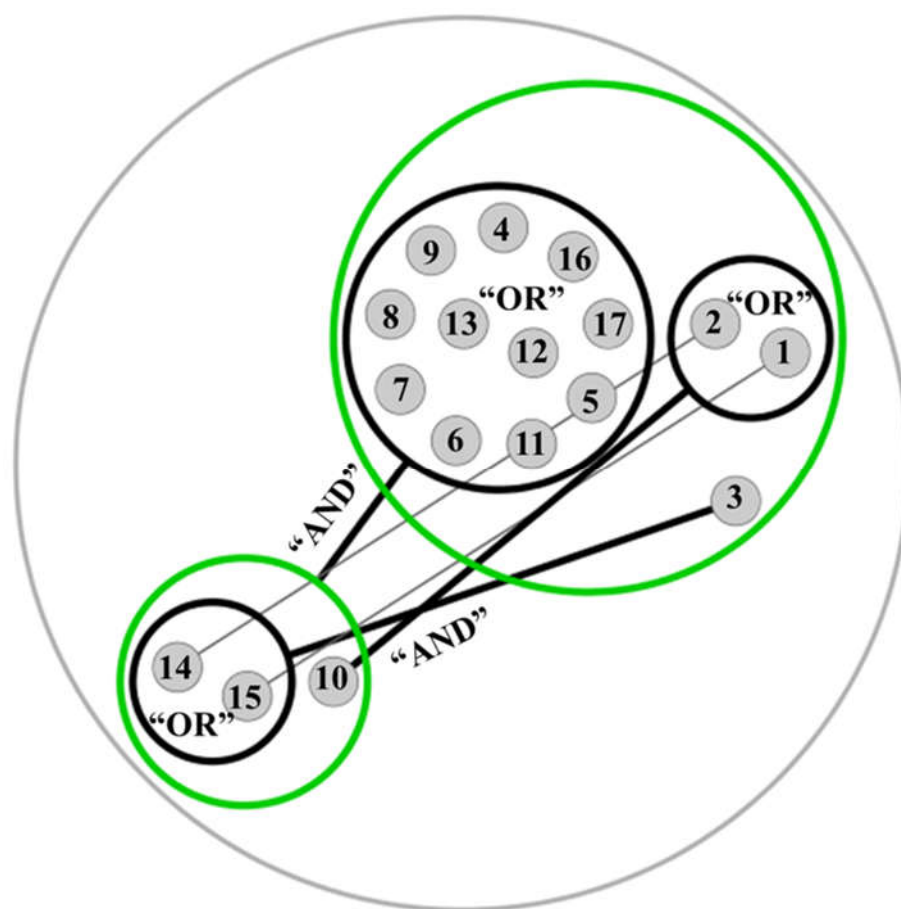


Figure S3. Metabolic network illustrating potential cross-feeding mechanisms between meta-genome-assembled genomes (MAGs) recovered from the microbiota of healthy individuals in the presence of arabinoxylan. These MAGs involve beneficial arabinoxylan and pectin degraders (*Akkermansia muciniphila*, *Barnesiella viscericola* DSM 18177, *Paraprevotella xylaniphila* YIT 11841) that were also found in the microbiota of patients with Crohn's disease (see Figure 1). In addition, MAGs of other microorganisms that were not identified in the microbiota of patients with Crohn's disease and might be greatly affected by gut dysbiosis were also chosen. Network nodes (i. e. circles containing different microbial communities showing equivalent metabolic functions) are connected by black lines indicating synergistic relationships between communities and complementary metabolic functions. Metabolic functions of MAGs from different nodes are needed to achieve the maximum number of end-products from arabinoxylan as well as other colonic metabolites (this mutualistic relationship is indicated by the conjunction "AND"). MAGs inside the same node play the same role and could be replaced by other members from the same community (this similar role is indicated by the conjunction "OR"). 1: *A. muciniphila*, 2: *B. viscericola* DSM 18177, 3: *P. xylaniphila* YIT 11841, 4: unidentified *Parabacteroides* species, 5: unidentified *Roseburia* species, 6: *Bifidobacterium adolescentis* ATCC 15703, 7: *Acidaminococcus fermentans* DSM 20731, 8: *Brachyspira murdochii* DSM 12563, 9: *Desulfovibrio piger*, 10: *Faecalitalea cylindroides* T2-87, 11: *Lactobacillus ruminis* ATCC 27782, 12: *Methanobrevibacter smithii* ATCC 35061, 13: *Streptococcus thermophilus*, 14: *Veillonella atypica*, 15: *Clostridioides difficile*, 16: *Lactobacillus mucosae* LM1, 17: *Veillonella parvula* HSIVP1.