


β -Glucan is a major growth substrate for human gut bacteria related to *Coprococcus eutactus*

Anna M. Alessi,^{1,2} Victoria Gray,^{1,3}
Freda M. Farquharson,¹ Adriana Flores-López,¹
Sophie Shaw,³ David Stead,¹ Udo Wegmann,^{2,4}
Claire Shearman,² Mike Gasson,²
Elaina S. R. Collie-Duguid,³ Harry J. Flint¹ and
Petra Louis ^{1*}

¹University of Aberdeen, Rowett Institute, Aberdeen, UK.

²Institute of Food Research, Norwich, UK.

³University of Aberdeen, Centre for Genome-Enabled
Biology and Medicine, Aberdeen, UK.

⁴School of Chemistry, University of East Anglia, Norwich
Research Park, Norwich, NR4 7TJ, UK.

Summary

A clone encoding carboxymethyl cellulase activity was isolated during functional screening of a human gut metagenomic library using *Lactococcus lactis* MG1363 as heterologous host. The insert carried a glycoside hydrolase family 9 (GH9) catalytic domain with sequence similarity to a gene from *Coprococcus eutactus* ART55/1. Genome surveys indicated a limited distribution of GH9 domains among dominant human colonic anaerobes. Genomes of *C. eutactus*-related strains harboured two GH9-encoding and four GH5-encoding genes, but the strains did not appear to degrade cellulose. Instead, they grew well on β -glucans and one of the strains also grew on galactomannan, galactan, glucomannan and starch. *Coprococcus comes* and *Coprococcus catus* strains did not harbour GH9 genes and were not able to grow on β -glucans. Gene expression and proteomic analysis of *C. eutactus* ART55/1 grown on cellobiose, β -glucan and lichenan revealed similar changes in expression in comparison to glucose. On β -glucan and lichenan only, one of the four GH5 genes was strongly upregulated. Growth on glucomannan led to a transcriptional response of many genes, in particular a strong upregulation of glycoside hydrolases involved in mannan degradation. Thus,

β -glucans are a major growth substrate for species related to *C. eutactus*, with glucomannan and galactans alternative substrates for some strains.

Introduction

Dietary fibre originates from plant cell wall polysaccharides including cellulose, hemicellulose and pectins (Flint *et al.*, 2012a). Their molecular structure is highly heterogeneous due to the presence of different monosaccharides, which are bound by a variety of glycosidic bonds. They are recalcitrant to digestion in the small intestine and reach the colon, where they serve as a substrate for microbial fermentation, which leads to the formation of short-chain fatty acids (SCFAs, mainly acetate, propionate and butyrate) and gases (Flint *et al.*, 2012a). Indigestible plant storage polysaccharides and oligosaccharides are also widely regarded as belonging to the fibre-fraction of foods, as they reach the large intestine intact (Howlett *et al.*, 2010; Slavin, 2013). An adequate supply of fibre and its efficient degradation is essential in fuelling the numerous health-promoting actions of the gut microbiota, in particular the provision of beneficial SCFAs (Flint *et al.*, 2012b; Louis and Flint, 2017). The fermentability of hemicelluloses, consisting of different types of polysaccharides (arabinoxylans, β -glucans, mannans, xyloglucans, etc.) is relatively high in the human gut and has been reported to rely on hydrolytic action of specific isolates belonging to many genera, including *Roseburia* and *Bacteroides* (Flint *et al.*, 2012a; Sheridan *et al.*, 2016; Tuncil *et al.*, 2017). In contrast, the breakdown of cellulose by the human gut microbiota is less efficient and appears to be restricted to few species (Cann *et al.*, 2016). *Ruminococcus champanellensis* is closely related to *Ruminococcus flavefaciens*, a major cellulose degrader in the herbivore GI tract (Flint *et al.*, 2008; Chassard *et al.*, 2011). Cellulolytic activity was also reported for a *Bacteroides* species isolated from the human gut (Robert *et al.*, 2007). There is evidence for inter-individual variation in cellulose-degrading gut microbes, which appears to correlate with whether methanogenic Archaea are present in an individual (Chassard *et al.*, 2010).

The colonic microbiota is highly complex and our knowledge to date of which microbes are instrumental in fibre breakdown is incomplete. Numerous carbohydrate-active

Received 3 July, 2019; revised 27 February, 2020; accepted 3 March, 2020. *For correspondence. E-mail p.louis@abdn.ac.uk; Tel. 44(+0) 1224 438735; Fax 44(+0) 1224 438699.

enzymes involved in fibre breakdown have been characterized and large human gut-derived data sets are available for genomic and metagenomic mining. However, it is often difficult to deduce function from sequence alone, as many glycoside hydrolase families comprise enzymes with different substrate specificities (Lombard *et al.*, 2014). Functional metagenomics, which relies on functional expression of environmental genes in a heterologous host, can be utilized to identify microbes that are involved in fibre breakdown in the gut and reveal novel enzymatic functions. This approach has successfully been applied to human gut microbiota (Tasse *et al.*, 2010; Cecchini *et al.*, 2013). The heterologous expression host commonly used is *Escherichia coli*, which has been reported to successfully express genes from a wide range of organisms (Handelsman, 2004). However, an *in silico* analysis of 32 prokaryotic genome sequences for the presence of expression signals functional in *E. coli* suggests that only approximately 40% of genes would be successfully expressed in this host, with extensive variation (7%–73%) between different organisms (Gabor *et al.*, 2004). Furthermore, post-translational processes, such as protein folding, insertion into the cell membrane or secretion from the cell may also differ between different microbes. The use of alternative expression hosts for metagenomic libraries may therefore improve the recovery of novel genes from metagenomic libraries, which has been demonstrated for functional metagenomic studies from other environments (McMahon *et al.*, 2012).

Here, we report the comparative analysis of *E. coli* XL1 Blue and *Lactococcus lactis* MG1363 as hosts for functional screening of a human faecal metagenomic library on a range of different dietary carbohydrates. *L. lactis* belongs to low %G + C Firmicutes, is well-characterized and widely used as an alternative host to *E. coli* for heterologous gene expression, and genetic tools and vectors are available (Pontes *et al.*, 2011). This led to the identification of a clone carrying a glycoside hydrolase (GH) family 9 gene with sequence identity to *Coprococcus eutactus* ART55/1. As this GH family is usually associated with cellulose breakdown, we investigated the breakdown of beta-linked glucans in this strain and its gene and protein expression response to growth on different substrates.

Results

Screening for glycoside hydrolase activities from a metagenomic library in *Escherichia coli* and *Lactococcus lactis*

A human gut microbiome metagenomic library (6146 clones, average insert size estimate 2.5 kb), constructed in shuttle vector pTRKL2 and transformed into *E. coli* XL1 Blue, was functionally screened for glycoside hydrolase (GH) activities using seven carbohydrate substrates. Enzyme activity was confirmed in 16 clones after re-streaking on the respective media. Positive clones were found for all substrates apart from polygalacturonic acid and rhamnopyranoside, with the

Table 1. Metagenomic clones with glycoside hydrolase activity detected by functional screening of clone libraries generated in *E. coli* XL1 Blue and *L. lactis* MG1363.

Clone ID	Length (nt)	Host	Enzyme activity ^a					Match organism	ORF matches with relevance for carbohydrate metabolism ^b
			AF	S	CMC	L	X		
P3H22	3203	<i>E. coli</i>	+					<i>Bacteroides plebeius</i>	Glycoside hydrolase, alpha-L-arabinofuranosidase, beta-galactosidase
P3B15	3470	<i>E. coli</i>	+++					<i>Eubacterium rectale</i>	Alpha-L-arabinofuranosidase, alpha-N-arabinofuranosidase
P1P9	5693	<i>E. coli</i>		+				<i>Bacteroides plebeius</i>	SusD family nutrient uptake, carbohydrate-binding protein
P5H21	4130	<i>E. coli</i>		+				<i>Bacteroides uniformis</i>	Type I pullulanase
P5E1	3949	<i>E. coli</i>		+	+/-			<i>Bacteroides</i> sp.	6-Phosphogluconate dehydrogenase
P5D24	5531	<i>E. coli</i>		+	+/-			<i>Alistipes putredinis</i>	
P3N11	3927	<i>E. coli</i>		+/-	+/-			Bacteroidales	
P1E14	3637	<i>E. coli</i>		+				<i>Eubacterium</i> sp.	Alpha amylase
P1116	2754	<i>E. coli</i>		++	+			<i>Faecalibacterium</i> sp.	
P3C3	2202	<i>E. coli</i>		++	+			Lachnospiraceae	Polysaccharide deacetylase
P1D18	2918	<i>E. coli</i>		+	+/-			<i>Anaerostipes hadrus</i>	
P8117	4094	<i>E. coli</i>		++	+			<i>Anaerostipes hadrus</i>	Glucosamine-6-phosphate deaminase, cell wall hydrolase
P5J3	4246	<i>E. coli</i>		++	+			Firmicutes	
P7I21	5402	<i>E. coli</i>		+				<i>Coprobacillus</i> sp.	
P5M23	2994	<i>E. coli</i>		+	+/-				
P9N7	2655	<i>E. coli</i>			+++	+++	++	<i>Alistipes senegalensis</i>	
P20A8	2643	<i>L. lactis</i>			+++	+++	++	<i>Coprococcus eutactus</i>	Bacterial surface proteins containing Ig-like domains

a. Substrates used for screening: AF, 4-methylumbelliferyl α -L-arabinofuranoside; S, potato starch; CMC, carboxymethyl cellulose; L, lichenan; X, oat spelt xylan. No clones with activity on polygalacturonic acid or 4-methylumbelliferyl α -L-rhamnopyranoside were detected. Level of activity detected is based on visual inspection of clearing zones on substrate-containing plates.

b. Full details of blast results of clone sequences are given in the Supporting Information Table S1.

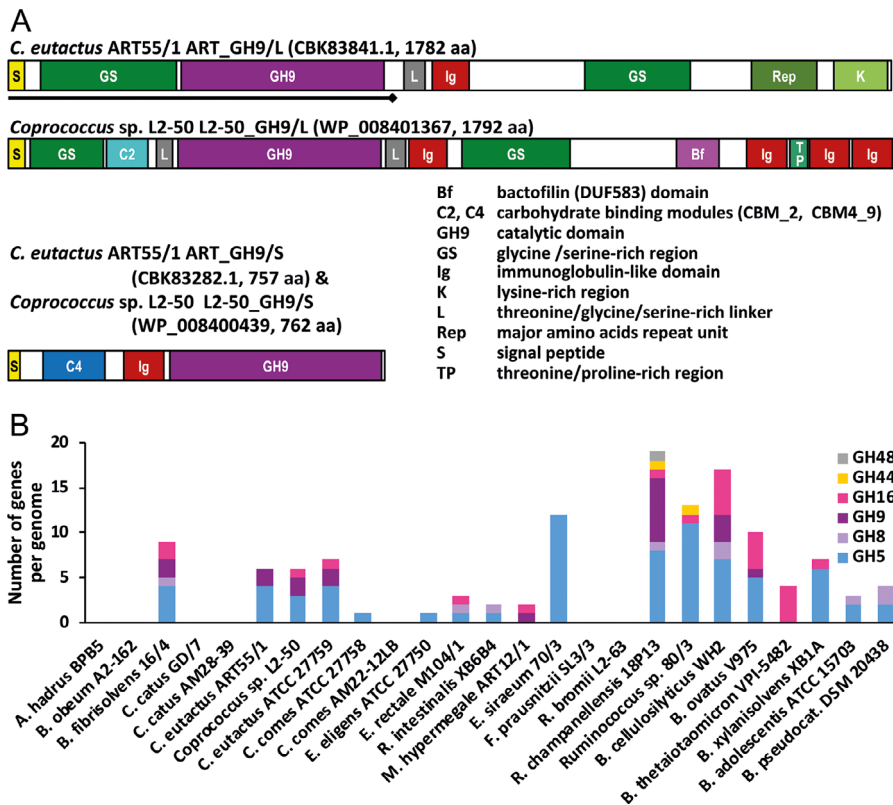


Fig. 1. Glycoside hydrolase genes associated with β -glucanase activity in human gut bacteria.

A. GH9-domain containing genes (short genes designated GH9/S and long genes GH9/L) in *C. eutactus* ART55/1 and *Coprococcus* sp. L2-50. Accession numbers and deduced length in amino acids (aa) are given in brackets. The start codon for L2-50_GH9/L was re-assigned to position 18 of WP_008401367.1 based on the presence of a ribosome-binding site (GGAAG, eight nucleotides upstream) and a signal peptide motif. The line below ART_GH9/L indicates the region covered by clone P20A8 from the metagenomic library. Domain structure predictions are based on PFAM, PROSITE, InterPro and SMART databases searching.

B. Genome carriage of glycoside hydrolase (GH) gene families associated with β -glucanase activity in human gut bacteria.

highest number of clones on starch and carboxymethyl cellulose (Table 1). The comparative analysis of most sequenced inserts from positive clones showed a high level of identity to sequences from a variety of Bacteroidetes and Firmicutes bacteria of gut origin (Table 1 and Supporting Information Table S1). Sequence analysis of open reading frames revealed homology with enzymes with the expected substrate specificity for some clones (e.g. P3H22 and P3B15, detected on α -L-arabinofuranoside; P5H21 and P1E14, detected on starch) whereas other clone sequences harboured less well-characterized open reading frames (Table 1 and Supporting Information Table S1).

The *E. coli* XL1 Blue library was pooled and transferred into *L. lactis* MG1363 (4608 clones, insert frequency estimate 75%, average insert size estimate 2.5 kb) and functionally screened on all seven carbohydrate substrates, which resulted in a total of three positive clones on the β -linked carbohydrates carboxymethyl cellulose (CMC), lichenan and xylan. Sequencing analysis revealed that all three clones contained identical insert sequences but were different from the single positive clone found on the same substrates in *E. coli* XL1 Blue (Table 1 and Supporting Information Table S1). One of the clones (P20A8) was transferred into *E. coli* XL1 Blue, but showed only very weak enzyme activity in this host. The 16 positive clones detected in *E. coli* XL1 Blue were also

transformed into *L. lactis* MG1363, but none of them displayed enzyme activity in this host (data not shown). Clone P20A8 showed the presence of a truncated ORF with 100% sequence identity to *C. eutactus* ART55/1 gene CBK83841.1 (Fig. 1A). It contained a GH9 catalytic domain, suggesting that it encodes a β -glucanase.

Distribution of β -glucanase gene families among human colonic bacteria

Following detection of the GH9 catalytic domain containing clone P20A8 from *C. eutactus* ART55/1, we performed in silico analysis of β -glucanase gene families among human colonic bacteria. Glycoside hydrolases that break down β -(1,4) linkages in glucan chains belong to multiple GH families, notably GH5, GH8, GH9, GH16, GH44 and GH48 (CAZy database at www.cazy.org; Lombard *et al.*, 2014). Figure 1B shows the distribution of the best characterized β -glucanase gene families across genomes available for selected human colonic bacteria from CAZy spanning Firmicutes, Bacteroidetes and Actinobacteria. Only GH5 is widely distributed (16/25 genomes) and this family is known to include enzymes with a very diverse range of specificities. In contrast, GH48 (generally encoding cellobiohydrolases) occurs only in the cellulolytic species *R. champanellensis* while GH44 (which has been implicated in xyloglucan utilization) is limited to two

species of *Ruminococcus*. The GH9 family, which includes many bacterial cellulases among cellulolytic bacteria from gut and non-gut habitats, also shows a limited distribution among human colonic anaerobes (Fig. 1B). This prompted us to investigate the characteristics of *Coprococcus*-related isolates and their GH9 genes further.

GH9 genes of *Coprococcus*-related gut isolates

Genome analysis of *C. eutactus* ART55/1 and a related strain, *Coprococcus* sp. L2-50 (96% identity of the respective 16S rRNA gene sequences AY350746 and AJ270491; Fig. 2), revealed that each organism harboured two genes containing GH9 catalytic domains (Fig. 1A). All four GH9 enzymes contained a typical N-terminal lipoprotein signal peptide and showed the presence of two conserved aspartate residues and a glutamate residue involved in the catalytic reaction (Kurokawa *et al.*, 2002; Pereira *et al.*, 2009; Pereira *et al.*, 2010). The shorter genes ART_GH9/S (CBK83282.1) and L2-50_GH9/S (WP_008400439.1) share a high level of identity (60%) and similar multi-domain structure. The catalytic domains of the longer proteins ART_GH9/L (CBK83841.1) and L2-50_GH9/L (WP_008401367.1) share 48% identity and the complete proteins exhibited a more diverse multi-domain architecture that differed between both enzymes (Fig. 1A).

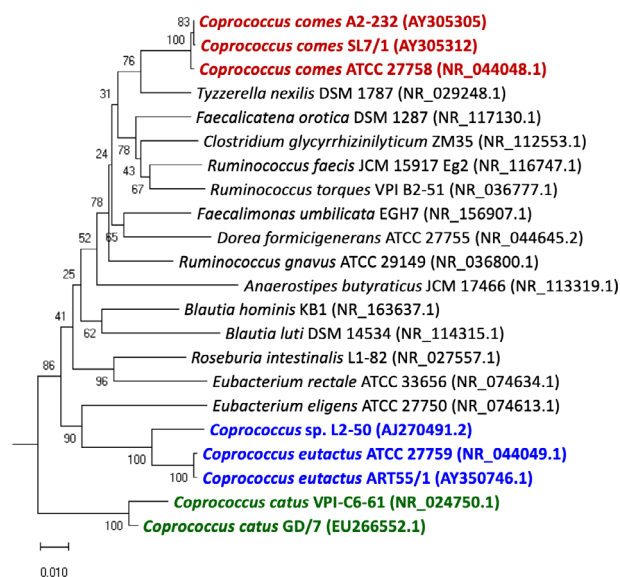


Fig. 2. Evolutionary relationships of *Coprococcus* species and their closest human gut bacteria relatives, inferred using the Neighbour-Joining method (Saitou and Nei, 1987). Bootstrap values (Felsenstein, 1985) from 500 replications are shown at branches and evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura *et al.*, 2004) using Mega X (Kumar *et al.*, 2018).

Phylogenetic analysis, based on multiple amino acid alignments of the catalytic domains, showed that ART_GH9/S and L2-50_GH9/S have >50% sequence identity to GH9 enzymes from other Clostridiales species including Lachnospiraceae bacteria, *Butyrivibrio* and *Ruminococcus* species within theme D of GH9 cellulases (Supporting Information Fig. S1). All of these enzymes display modular architecture similar to ART_GH9/S and L2-50_GH9/S with the presence of a carbohydrate-binding domain (CBM4_9, pfam02018), Ig-like domain (cd02850) and glycosyl hydrolase family 9 catalytic domain (pfam00759). The amino acid sequences of the GH9/L catalytic modules from both strains showed a lower level of identity (40%–50%) to other Clostridiales species belonging to Lachnospiraceae and *Ruminococcus* (Supporting Information Fig. S1).

The genus *Coprococcus* within the Lachnospiraceae family of Firmicutes contains three species, however, they are not phylogenetically closely related (Fig. 2). Sequence similarity searches of both GH9 genes identified in *C. eutactus* ART55/1 against reference genomes of *C. eutactus*, *Coprococcus catus* and *Coprococcus comes* revealed that both genes were present in eight *C. eutactus* strains including the type strain *C. eutactus* ATCC 27759 (ART_GH9/S, CBK83282.1, query coverage >91%, sequence identity >75%; ART_GH9/L, CBK83841.1, query coverage >97%, sequence identity >68%). No significant similarity was found with either GH9 gene in the *C. catus* and *C. comes* genomes. CAZyme analysis of several genomes of different *Coprococcus* species further confirmed that the GH9 genes were only present in *C. eutactus*, whereas in the other species, no β -glucanase-related genes were found apart from a single GH5 gene in one of the examined *C. comes* strains (Fig. 1B).

Catalytic activities of the GH family 9 enzymes from *C. eutactus* ART55/1 and *Coprococcus* sp. L2-50

The full length GH9 genes including predicted promoter sites were cloned in the same orientation into shuttle vector pTRKL2 and transformed into *E. coli* XL1 Blue and *L. lactis* MG1363. The transformants exhibited enzyme activity on substrate-containing plates using the Congo Red detection method (Fig. 3A). The enzymes from *Coprococcus* sp. L2-50 were functionally expressed in both hosts and showed activity on CMC- and lichenan-containing plates. In contrast, the enzymes from *C. eutactus* ART55/1 were functionally expressed only in *L. lactis* MG1363. The *E. coli* XL1 Blue transformants were devoid of lichenase activity and showed only limited activity on CMC-containing plates (Fig. 3A). Analysis of codon usage of the two GH9 genes from *C. eutactus* ART55/1 revealed that differences in codon usage may be

responsible for their poor expression in *E. coli* XL1 Blue, as two codons of infrequent use in *E. coli* (Zhang *et al.*, 1991) were present at relatively high frequency (ART_GH9/S AGA 14.5/1000, AUA 25.1/1000; ART_GH9/L AGA 9.5/1000, AUA 26.9/1000). The *Coprococcus* sp. L2-50 genes, on the other hand, had mostly lower frequencies (L2-50_GH9/S AGA 11.8/1000, AUA 0/1000; L2-50_GH9/L AGA 6.6/1000, AUA 7.7/1000) in line with their higher expression levels in *E. coli* XL1 Blue. The enzyme catalytic activities were also examined in supernatants and cell-free extracts of grown cultures using a reducing sugar assay, which agreed with the plate assay results. Enzyme activities for *E. coli* XL1 Blue transformants were mainly associated with the cell-free extracts, whereas activities from *L. lactis* MG1363 cultures were mainly detected in the supernatant fraction (Supporting Information Table S2). Thus, the cloned enzymes appear to be secreted by *L. lactis* MG1363.

To investigate the functionality of the different domains present in the shorter GH9 gene that was strongly conserved between both strains (Fig. 1A), different constructs that lacked either the carbohydrate-binding domain and/or the Ig-like domain were generated of L2-50_GH9/S (Fig. 3B). The constructs were cloned and overexpressed in *E. coli* BL21 (DE3), and enzyme activities on various substrates were determined. Enzyme activities of the construct lacking the CBM did not differ from the complete enzyme, but the deletion of the Ig-like

domain abolished catalytic activity (Fig. 3B). The highest activity was found on β -glucan, followed by lichenan.

Growth of *Coprococcus* species on different carbohydrates

We examined the ability of *C. eutactus* ART55/1 and *Coprococcus* sp. L2-50 to utilize a wide range of carbohydrate substrates by measuring optical density and a drop in pH. *Coprococcus* sp. L2-50 exhibited growth on glucose, cellobiose, β -glucan and lichenan as well as very limited growth on potato starch, but no utilization could be detected on laminarin, glucomannan, galactomannan, mannan, galactan, xylan, xyloglucan, arabinoxylan and pullulan (Fig. 4A and Supporting Information Fig. S2). In addition to the substrates utilized by *Coprococcus* sp. L2-50, *C. eutactus* ART55/1 was also able to grow on glucomannan, galactomannan, galactan and potato starch and a limited pH drop was also detected on mannan (Fig. 4 and Supporting Information Fig. S2). Medium pH was tracked for up to 11 days on the insoluble substrates Sigmacell type 50, acid-swollen cellulose and filter paper for both *C. eutactus*-related strains. No decrease in pH was seen (data not shown), indicating a failure to utilize these substrates. Thus, both strains are non-cellulolytic but are able to utilize certain soluble β -glucans, suggesting a possible role for the GH9 enzymes given their activity against these substrates (Fig. 3). Substrate utilization

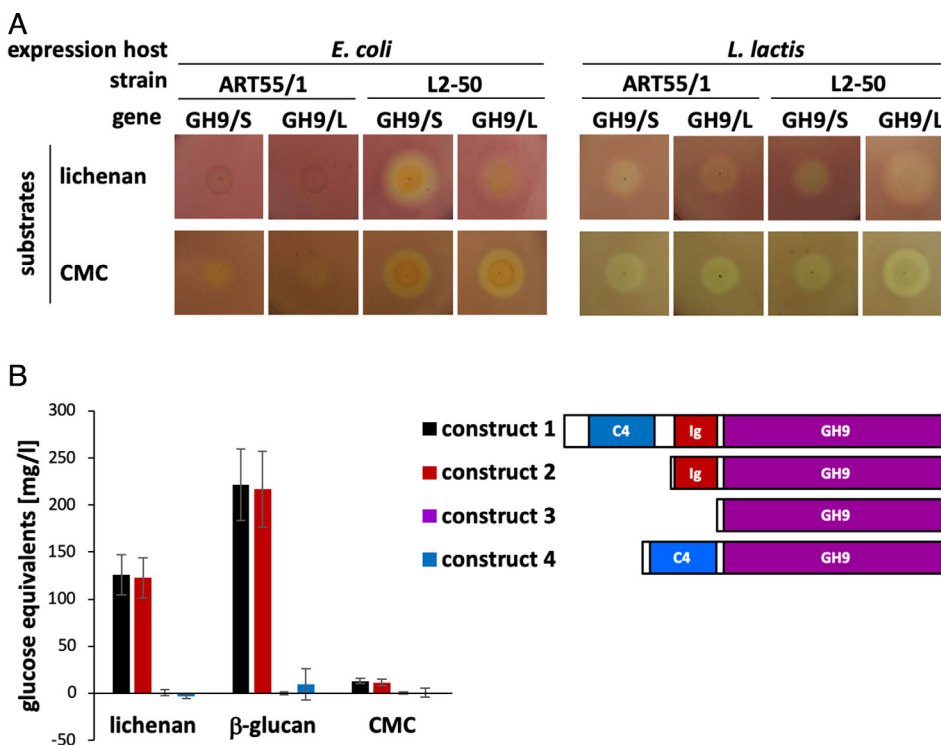


Fig. 3. Enzyme activities of GH9 genes from *Coprococcus*-related strains.

A. Fibrolytic activity of *E. coli* and *L. lactis* transformants with *C. eutactus* ART55/1 and *Coprococcus* sp. L2-50 GH9 genes on lichenan- or CMC-containing agar plates after overnight incubation of 10 μ l of a freshly grown overnight culture and staining with Congo Red.

B. Enzyme activity of different constructs of *Coprococcus* sp. L2-50 L2-50_GH9/S missing either the carbohydrate binding and/or Ig-like domain (for domain designations see Fig. 1). Constructs were cloned and overexpressed in *E. coli* BL21 (DE3). Release of reducing sugars from lichenan, β -glucan and CMC was determined by Lever assay over 1 h of incubation and background activity of *E. coli* BL21 (DE3) carrying the empty vector were subtracted. Mean and standard deviation of three replicates.

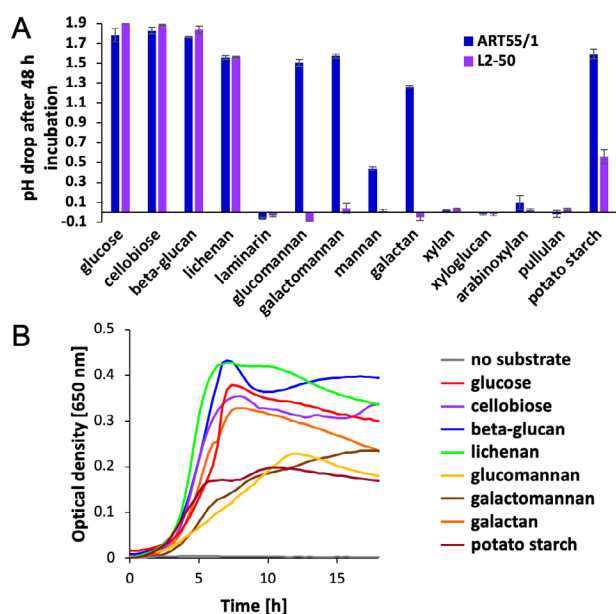


Fig. 4. Carbohydrate utilization of *C. eutactus* ART55/1 and *Coprococcus* sp. L2-50 on different substrates.

A. Medium pH drop after growth in 96-well plates after 48 h of incubation (mean and standard deviation of triplicate cultures). Growth on cellulose was carried out in Hungate tubes to enable longer incubation periods, but the pH did not change after 11 days of incubation. B. Representative growth curves of *C. eutactus* ART55/1 during growth in 96-well plates. No or very little growth was observed on laminarin, xylan, xyloglucan, arabinoxylan and pullulan (OD <0.07).

was also examined of further strains from the three different *Coprococcus* species, which revealed that *C. eutactus* ATCC 27759 showed a very similar behaviour to *C. eutactus* ART55/1, but the three *C. comes* strains showed good growth only on glucose and *C. catus* GD/7 only showed very limited growth on potato starch (Supporting Information Fig. S2). This is in agreement with both the phylogenetic placement of the species as well as with the absence of the respective glycoside hydrolase genes in those strains.

Induction of β -glucanase activity during growth on β -glucan and transcriptional response to growth on different carbohydrates in *C. eutactus* ART55/1

Coprococcus eutactus ART55/1 and *Coprococcus* sp. L2-50 were grown to early stationary phase on either glucose, cellobiose or β -glucan to investigate whether β -glucanase activity was inducible in these strains. β -glucanase activity was much higher for cell extracts and supernatants from cultures grown on β -glucan than on glucose or cellobiose, especially for *C. eutactus* ART55/1 (Supporting Information Table S3). To identify differentially expressed genes (DEGs) during growth on β -glucan-type carbohydrates, *C. eutactus* ART55/1 grown on glucose was inoculated into medium

containing glucose, cellobiose, β -glucan, lichenan or glucomannan and grown to exponential phase (Supporting Information Table S4) for RNA extraction and transcriptomic analysis by RNA sequencing. Principal component analysis (PCA) of transcript changes showed that *C. eutactus* ART55/1 cultures grown on glucomannan clustered most distinctly, whereas more similarity was seen between cellobiose, β -glucan and lichenan incubations (Fig. 5A). Significant DEGs [false discovery rate (FDR) <0.05 and log-fold change (LogFC) >1] compared to glucose as baseline were determined for all carbon sources. Glucomannan showed 197 DEGs compared to glucose specific to this substrate, whereas the other three substrates each only exhibited between two and seven specific DEGs (Fig. 5B). For all DEGs per substrate (including those shared with other substrates), on glucomannan almost half of the genes were upregulated (44%) relative to glucose, whereas on the other three substrates almost three quarters of the genes were downregulated (cellobiose 71%, β -glucan 72%, lichenan 68%, Fig. 5B). This trend was also reflected in the magnitude of the response (maximum fold change of upregulated DEGs 3.6, 20.7, 16.7 and 560.7 on cellobiose, β -glucan, lichenan and glucomannan, respectively, downregulated DEGs 273.8, 295.0, 231.9 and 40.6, respectively, Supporting Information Table S5).

Thirty-eight genes were significant across all four comparisons, with an additional 25 shared between β -glucan, cellobiose and lichenan (Fig. 5B). Interestingly, the complete data set of all 299 DEGs showed a very strong positive correlation for all pairwise comparisons between cellobiose, β -glucan and lichenan ($P < 0.0001$) but not for comparisons with glucomannan (Fig. 5C and Supporting Information Fig. S3). Many of the DEGs significant on all four substrates responded in the opposite direction on glucomannan relative to each of the other substrates (Fig. 5C and Supporting Information Fig. S3, data labelled in black). Gene ontology (GO) enrichment analysis identified several GO terms that were significantly over represented in the DEGs. These were mostly related to localization, transport, carbohydrate metabolic processes and hydrolase activity (Supporting Information Table S6). Three of the four genes strongly downregulated on all four substrates code for a CUT1 family ABC transporter with the fourth gene encoding a GH77 (4- α -glucanotransferase). On β -glucan, lichenan and to a lesser degree cellobiose, another ABC transporter belonging to the CUT2 family was also significantly downregulated. On glucomannan, on the other hand, an ABC-transporter annotated as a multidrug transport system was strongly upregulated (Supporting Information Table S5).

The genome of *C. eutactus* ART55/1 contains 45 glycoside hydrolase genes according to the CAZy database (Supporting Information Table S7), 19 of which showed a

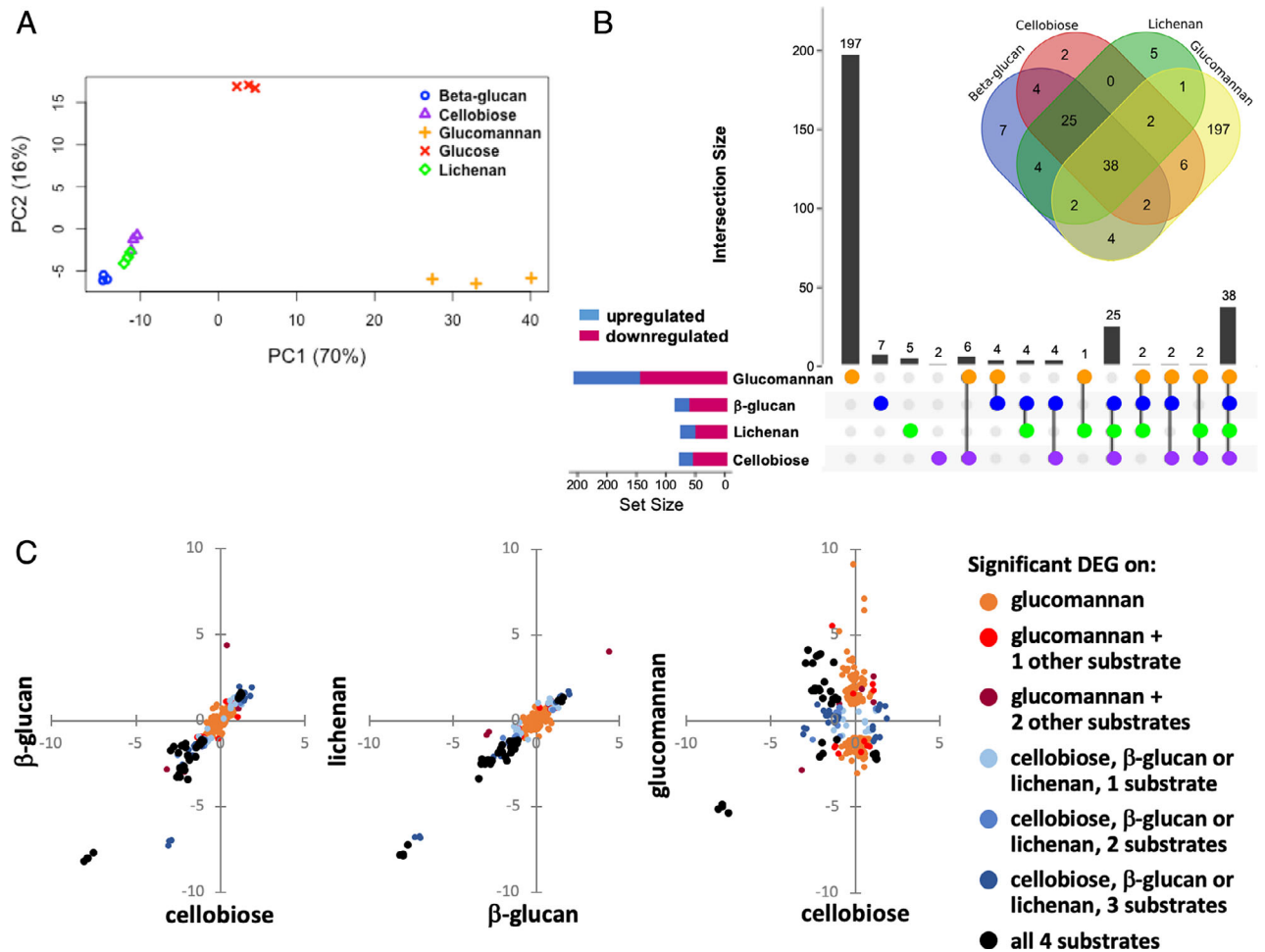


Fig. 5. Gene expression changes after growth of *C. eutactus* ART55/1 on five different substrates. A. PCA showing separation of the samples based on differences in gene expression with samples coloured by carbon source. B. UpSet plot and Venn diagram summarizing total number of genes differentially expressed compared to glucose, direction of change and intersections between sets of differentially expressed genes. C. Relationship of all differentially expressed genes (DEGs, 299 genes in total) between different growth substrates, expressed as logFC. The remaining comparisons are shown in Fig. S3.

significant change in gene expression (Fig. 6). One of the four GH5 genes present in the genome was strongly upregulated on β -glucan and lichenan and to a lesser degree on glucomannan. One of the two GH9 genes (GH9/L) was significantly upregulated on all four substrates, with glucomannan showing the strongest response. For most other GH genes, the response was opposite for glucomannan compared to the other substrates, with the vast majority being upregulated on glucomannan. The strongest upregulation was observed for all five genes assigned to GH families predominantly involved in mannan degradation (GH26, GH113 and GH130; Fig. 6). All ribosomal proteins were significantly reduced (FDR <0.05) on glucomannan and several reached a logFC of over -1 , whereas none of the other substrates showed any significant changes relative to glucose (Supporting Information Table S5). This is in

agreement with the lower growth rate observed on glucomannan (Supporting Information Table S4).

Proteomic response to growth on different carbohydrates in C. eutactus ART55/1

Bacterial cells from the cultures used for gene expression analysis were also subjected to proteomics analysis by mass spectrometry. In total, 891 *C. eutactus* ART55/1 proteins were identified. Label-free quantification (LFQ) values were analysed for differential abundance with a linear model identical to the gene expression analysis with glucose as the baseline and all proteins with adjusted significance values below 0.05 were regarded as differentially abundant. PCA revealed strikingly similar relationships between the different samples as was found

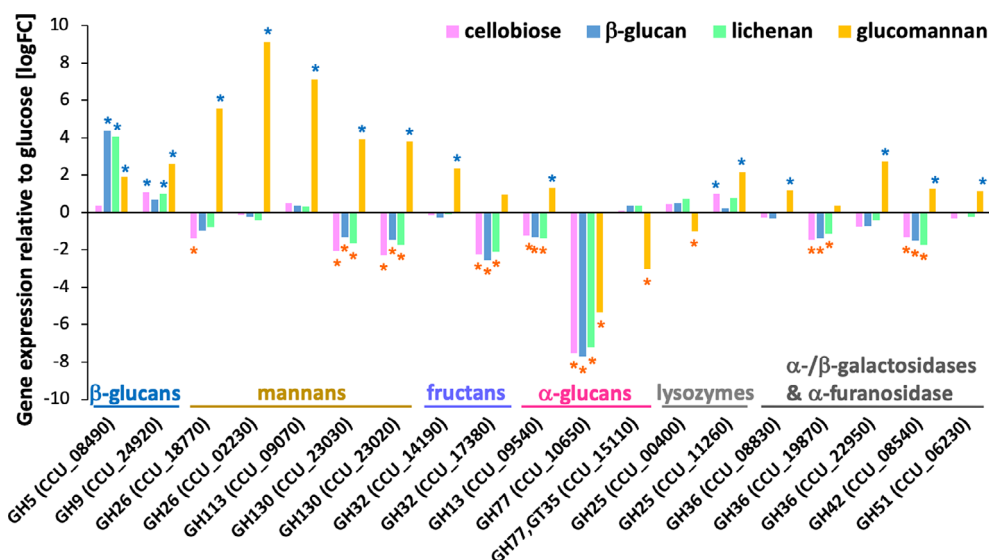


Fig. 6. Carbohydrate-active genes differentially expressed compared to glucose in *C. eutactus* ART55/1 during growth on cellobiose, β -glucan, lichenan and glucomannan.

Putative enzyme substrates based on CAZypedia (CAZypedia Consortium, 2018) are indicated above the GH genes. Significant up- and down-regulation for each of the four growth substrates relative to glucose is indicated by blue and orange asterisks, respectively.

for the gene expression data (Fig. 7A). A comparison of LFQ values with reads per kilobase per million mapped reads (RPKM) values from the gene expression analysis showed a strong positive trend between the two datasets (Fig. 7B) despite the fact that they showed only a partial overlap in significantly differentially expressed/abundant genes and proteins (Supporting Information Fig. S4), likely reflecting post-transcriptional networks regulating protein expression (Vogel and Marcotte, 2012). As for the gene expression data, glucomannan resulted in the largest difference in the proteome compared to glucose, and the overall distribution of shared proteins between different substrates was similar as well (Fig. 7C). Several expression changes identified at the transcript level were confirmed at the proteome level, including the upregulated GH5 and some of the proteins comprising the downregulated CUT1 and CUT2 ABC transporters on β -glucan-type substrates, as well as some of the carbohydrate active enzymes upregulated on glucomannan (Supporting Information Table S5).

Discussion

Functional metagenomic screening was employed here to identify genes involved in carbohydrate breakdown from human faecal microbiota. A comparison of the suitability of two cloning hosts, *E. coli* XL1 Blue and *L. lactis* MG1363, revealed large differences in their functional expression of heterologous genes. Several active clones from a wide range of bacteria within both the Bacteroidetes and Firmicutes were recovered from the *E. coli* XL1 Blue

library, whereas only a single clone with β -glucanase activity was recovered after electroporation of the *E. coli* XL1 Blue library into *L. lactis* MG1363. Thus, *L. lactis* MG1363 appears to be quite limited in its ability to successfully express genes from other organisms, and alternative Gram-positive hosts (e.g. *Bacillus subtilis*, Dobrijevic *et al.*, 2013) may be more suitable for future functional metagenomic studies. This study does, however, also show the limitations in using *E. coli* as heterologous host, as the clone expressing very high activity in *L. lactis* MG1363 (P20A8) did not exhibit good activity in *E. coli* XL1 Blue. *E. coli* has been estimated to express approximately 40% of enzymatic activities from diverse microbial origins based on the analysis of expression signals in microbial genomes and is strongly biased towards genes from certain groups of organisms (Uchiyama and Miyazaki, 2009). Differences in codon usage between *C. eutactus* ART55/1 and *E. coli* XL1 Blue likely contribute to the poor expression of the two GH9 genes in *E. coli*. Protein export from the cell can also significantly affect successful heterologous expression (Freudl, 2018). Fractionation of grown cultures and enzyme activity measurements showed that *E. coli* XL1 Blue was not able to export the extracellular enzymes efficiently, whereas the activity was mainly detected in the culture supernatant in *L. lactis* MG1363. Differences in signal peptide recognition, secretion mechanism, the presence of chaperones and architecture of the cell envelope can affect secretion abilities between different organisms (Mingardon *et al.*, 2011; Burdette *et al.*, 2018).

The glycoside hydrolase family 9 mainly consists of cellulases (Wilson and Urbanowicz, 2019) and GH9 genes

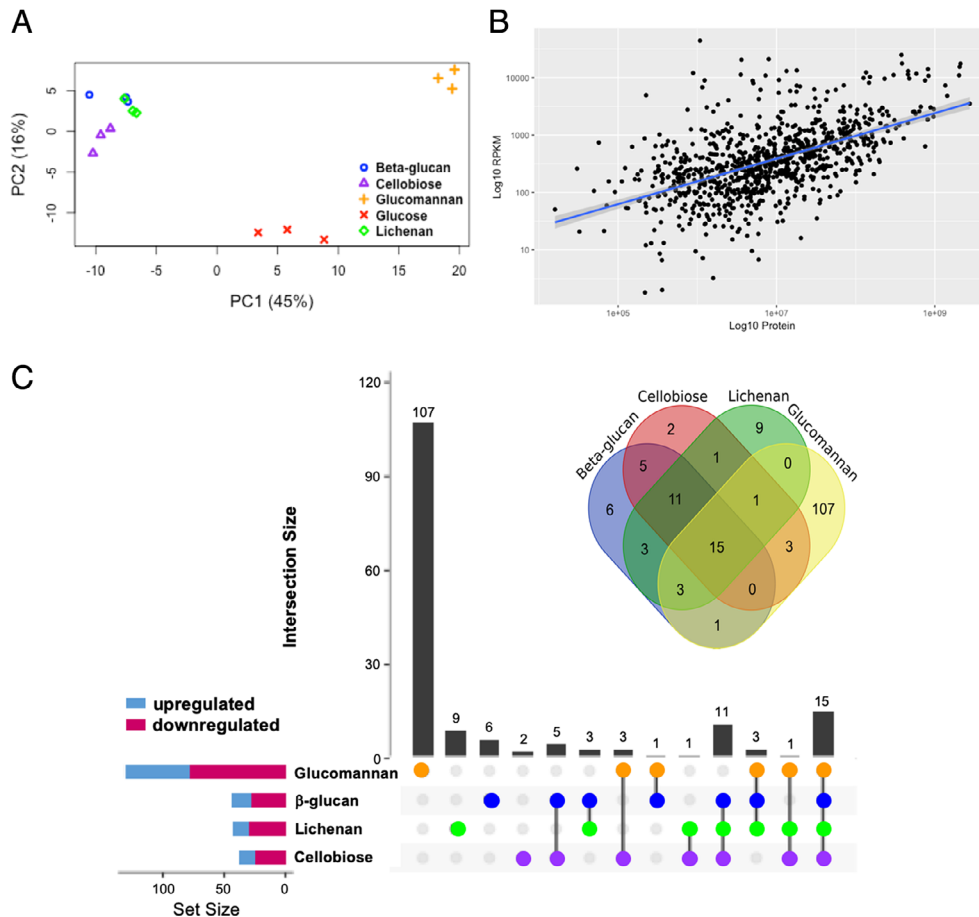


Fig. 7. Proteomic changes after growth of *C. eutactus* ART55/1 on five different substrates.

A. PCA showing separation of the samples based on differences in protein expression with samples coloured by carbon source.

B. Correlation between transcript and proteome expression levels for all detected proteins (log₁₀ average protein expression versus log₁₀ average RPKM for each gene).

C. UpSet plot and Venn diagram summarizing total number of proteins differentially expressed compared to glucose, direction of change and intersections between sets of differentially expressed genes.

in the human cellulose degrader *R. champanellensis* have been shown to be involved in cellulose breakdown (Morais *et al.*, 2016). This prompted us to investigate whether the *C. eutactus*-related strains ART55/1 and L2-50 are able to grow on cellulose, but this was not the case. Instead, both strains showed excellent growth on barley β -glucan and lichenan. The carriage of GH9 genes is therefore not a clear indicator of cellulose-degrading capacity in human gut bacteria. Based on sequence similarity and domain structure, ART_GH9/S and L2-50_GH9/S belong to GH9 theme D. Enzymes belonging to this group have been shown to initially cleave cellulose in a random mode and then act mainly as cellobiohydrolases (Devillard *et al.*, 2004). The catalytic domains of ART_GH9/L and L2-50_GH9/L show sequence similarity with *R. champanellensis* GH9A, which was classed as an endoglucanase (Morais *et al.*, 2016). More generally, GH9 enzymes have been associated with eight different catalytic specificities (EC numbers) however and these include activity against mixed-linkage beta

glucans, as detected here (CAZY website www.cazy.org, Lombard *et al.* 2014). *Coprococcus* sp. L2-50 did not exhibit good growth on any of the other eight polysaccharides tested, but *C. eutactus* ART55/1 and ATCC 27759 showed some growth on glucomannan, galactomannan, galactan and starch. Genome analysis revealed that *Coprococcus* sp. L2-50 contains significantly fewer glycoside hydrolase genes [30 based on database dbCAN2 (Zhang *et al.*, 2018)] than *C. eutactus* ART55/1 (43 based on dbCAN2, 45 based on CAZy, www.cazy.org) (Supporting Information Table S7). In agreement with the differences in carbohydrate degradation capacity seen for the two strains (Fig. 4), no genes belonging to GH26, GH113 or GH130, that were highly upregulated in *C. eutactus* ART55/1 on glucomannan, were identified in *Coprococcus* sp. L2-50 (Supporting Information Table S7). Despite the inability of *Coprococcus* sp. L2-50 to degrade galactan, the genome carriage of putative β -galactanases (CAZy Consortium, 2018) was similar between the two strains (one GH16 in *Coprococcus*

sp. L2-50 and one GH53 in *C. eutactus* ART55/1, Supporting Information Table S7). *Coprococcus* sp. L2-50, however, harboured only a single potential β -galactosidase (belonging to GH2), whereas *C. eutactus* ART55/1 encoded four (two GH1, one GH2 and one GH42). GH42 enzymes have been hypothesized to be involved in plant cell wall degradation and may work in cooperation with GH53 galactanases (Moracci, 2019).

Gene expression and proteomic analysis of *C. eutactus* ART55/1 on β -glucan-type substrates compared to glucose revealed a significant increase in one of the two GH9 genes, which was strongest on glucomannan. GH5 enzymes also hydrolyse β -glucan-type linkages and four GH5 genes are present in this strain. One of these GH5 genes was found here to be strongly upregulated, in particular on β -glucan and lichenan, making it a strong candidate for involvement in their degradation. The bioinformatic analysis showed that this GH5 protein (CCU_08490) belongs to subfamily 37, which are intracellular enzymes of bacterial origin with endo- β -1,3/4-glycanase (EC 3.2.1.4 and EC 3.2.1.73) and cellodextrinase (EC 3.2.1.74) activities (Aspeborg *et al.*, 2012). Interestingly, most GH enzymes in both strains appear not to be secreted via typical secretion systems, as they contain no predicted signal peptides (signal peptides detected in four of 30 enzymes in *Coprococcus* sp. L2-50 and seven of 45 enzymes in *C. eutactus* ART55/1, Supporting Information Table S7). The absence of predicted signal peptides on the majority of GH enzymes was previously observed in other members of the Lachnospiraceae family (Sheridan *et al.*, 2016). It remains to be established whether some of the GH enzymes without predicted signal peptides are membrane-associated or contain atypical secretory signal peptides (Gagic *et al.*, 2016); however, the main ecological niche of the *C. eutactus*-related strains appears to be in the breakdown of soluble and shorter length β -glucans rather than complex insoluble fibre.

The relative expression levels during growth on glucose (percentage of specific gene relative to all genes, data not shown) did not reveal a big difference between the six GH5 and GH9 genes (0.011%–0.022% for the four GH5 genes (with the lowest one upregulated), 0.026 and 0.037 for the two GH9 genes (with the higher one upregulated), overall range of all genes 0.000035%–2.55%). Therefore, differences in responses to growth on polysaccharides are likely not due to differences in basal gene expression between the GH5 and GH9 genes. The glycoside hydrolases strongly upregulated on glucomannan (five genes belonging to GH26, GH113 and GH130, see Fig. 6) on the other hand had a lower basal gene expression on glucose (0.0006%–0.0093%). In general, the overall gene expression profile was very similar on cellobiose, β -glucan and lichenan and the response in comparison to glucose

tended towards downregulation, with particularly strong responses for transporters likely involved in glucose transport. Thus β -glucans may be the preferred substrates for this organism, and oligosaccharides may be transported into the cell rather than glucose during growth on these substrates. The strong upregulation of genes involved in glucomannan degradation on this substrate, on the other hand, indicates that glucomannan constitutes an alternative energy source for *C. eutactus* ART55/1. *Coprococcus* sp. L2-50, a close relative of *C. eutactus* ART55/1, did not show good growth on any of the non- β -glucan-type substrates tested here, suggesting that bacteria related to *C. eutactus* may be adapted to thrive on β -glucans in the human large intestine. The negative correlation of GH77 on β -glucan, cellobiose, lichenan and particularly glucomannan in our study suggests that availability of these substrates repressed 4- α -glucanotransferase activity in *C. eutactus* ART55/1 associated with starch utilization (Ze *et al.*, 2015). The genus *Coprococcus* was increased on a high resistant starch diet in pigs and its abundance was positively correlated with starch breakdown products (Sun *et al.*, 2016). Both *C. eutactus* ART55/1 and *Coprococcus* sp. L2-50 carry several GH13 genes (Supporting Information Table S7), encoding α -amylases (Ze *et al.*, 2015), however, reasonably good growth on potato starch was only found for *C. eutactus* ART55/1 (and also for the second *C. eutactus* strain ATCC 27759), but final optical densities were low compared to β -glucans (Supporting Information Fig. S2). The three *C. comes* strains and *C. catus* GD/7, on the other hand, did not show good growth on any of the polysaccharides examined here and thus occupy different ecological niches from *C. eutactus*. They may degrade polysaccharides not included here or may cross-feed from breakdown products of primary polysaccharide degraders. *C. catus* grows well on fructose, but also grows on lactate, thus it is able to cross-feed on fermentation products from other bacteria (Reichardt *et al.*, 2014). Both species will have to be assigned new genus names based on their phylogenetic placement as well as physiological characteristics, as *C. eutactus* is the type species of the genus *Coprococcus* (Holdeman and Moore, 1974). The evidence provided here should also be taken into consideration for the interpretation of sequence-based studies, which often do not resolve data beyond genus level. Sequence-based studies that find a change in *Coprococcus* spp. should ideally be followed up with species-specific methods such as qPCR (Reichardt *et al.*, 2018).

Stimulation of *C. eutactus* may have beneficial effects on human health, as it contributes to the production of the health-promoting metabolite butyrate (Louis *et al.*, 2004). *Coprococcus* spp. were also consistently associated with higher quality of life across several cohorts and depleted in depression (Valles-Colomer *et al.*, 2019) and

C. eutactus showed an increase in abundance with a decrease in atopic dermatitis in an infant cohort (Nylund *et al.*, 2015). Our study will aid in the development of nutritional strategies to stimulate these potentially beneficial microbes in the gut.

Experimental procedures

Bacterial strains, plasmids and growth conditions

For the metagenomic library construction, functional screening and gene cloning *Escherichia coli* XL1 Blue (Stratagene, Milton Keynes, UK) and *Lactococcus lactis* MG1363 (Gasson, 1983) were used. *E. coli* XL1 Blue clones containing plasmid pTRKL2 (O'Sullivan and Klaenhammer, 1993) were selected on BHI (#CM1136; Oxoid, Basingstoke, UK) agar plates supplemented with 150 µg.ml⁻¹ erythromycin. *L. lactis* MG1363 clones containing plasmid pTRKL2 were selected on GM17 (#CM0817; Oxoid, supplemented with 0.5% (w/v) glucose) agar plates supplemented with 5 µg.ml⁻¹ erythromycin. *E. coli* XL1 was grown at 37°C with shaking at 220 rpm, and *L. lactis* MG1363 was grown at 30°C under static conditions.

For overexpression of synthetic gene constructs of *Coprococcus* sp. L2-50 gene GH9/S (WP_008400439), *Escherichia coli* BL21(DE3) was grown in Luria-Bertani broth (Sigma, Dorset, UK) with 50 µg.ml⁻¹ ampicillin at 37°C, with orbital shaking at 180 rpm.

Human faecal isolates *C. eutactus* ART55/1 (Louis *et al.*, 2004; note that its Genbank designation is *Coprococcus* sp. ART55/1 but for clarity it is named based on its phylogeny here as per Fig. 2), *Coprococcus* sp. L2-50 (Barcenilla *et al.*, 2000; note that its Genbank designation is *Clostridium* sp. ART55/1 but for clarity it is named based on its phylogeny here as per Fig. 2), *C. eutactus* ATCC 27759, *C. comes* ATCC 27758, *C. comes* A2-232 (Barcenilla, 1999), *C. comes* SL7/1 (Louis *et al.*, 2004) and *C. catus* GD/7 (Reichardt *et al.*, 2014) were maintained anaerobically on M2GSC medium at 37°C (Miyazaki *et al.*, 1997). Growth tests on different carbohydrates were performed in modified yeast extract-casitone-fatty acids (YCFA) medium (Duncan *et al.*, 2002) in 96-well plates in an anaerobic cabinet at 37°C under 10% (v/v) carbon dioxide, 10% (v/v) hydrogen and 80% (v/v) nitrogen atmosphere with a medium pH at start of the experiment of 6.5 ± 0.2. Bacterial cellulase activity was determined using cultures grown in Hungate tubes with different cellulosic substrates. Optical density and pH readings were used as indicators for growth. For transcriptomic and proteomic analysis, overnight cultures grown in YCFA medium containing 0.2% (w/v) glucose were inoculated into YCFA containing 0.2% (w/v) of one of the following carbohydrates: glucose, cellobiose,

barley β-glucan, lichenan or glucomannan and grown to exponential phase. Full details on media and growth conditions are given in supplemental methods.

Plasmid metagenomic library construction, functional screening and bioinformatic analysis

A freshly voided faecal sample was collected from a healthy female volunteer who had not received any antibiotics or other drugs during 6 months prior the sampling. The detailed method of library construction and functional screening is given in supplemental methods. Briefly, total metagenomic DNA was extracted, size fractionated and cloned into shuttle plasmid pTRKL2 (O'Sullivan and Klaenhammer, 1993). Transformation into *E. coli* XL1 Blue and selection of white colonies resulted in a library of 6146 viable clones containing an insert with an average insert size of 2.5 kb based on PCR colony screening of 24 randomly picked colonies. The library was transferred by pooling the *E. coli* XL1 Blue library and transforming the extracted DNA into *L. lactis* MG1363. It consisted of 4608 clones with 75% insert frequency and an average insert size estimated at 2.5 kb based on PCR colony screening of 15 random clones. Both libraries were arrayed on agar plates containing BHI (*E. coli*) or GM17 (*L. lactis*) medium with potato starch (Sigma-Aldrich S2004, 1% w/v), carboxymethyl cellulose (CMC, Sigma-Aldrich C4888, 0.5% w/v), lichenan (Sigma-Aldrich L6133, 0.05% w/v), oat spelt xylan (Sigma-Aldrich X0627, 0.5% w/v), polygalacturonic acid (Sigma-Aldrich P0853, 0.5% w/v), 4-methylumbelliferyl α-L-arabinofuranoside (Sigma-Aldrich M9519, 50 µg/ml) or 4-methylumbelliferyl α-L-rhamnopyranoside (Sigma-Aldrich M8412, 50 µg/ml) and incubated overnight at 37°C (*E. coli*) or 30°C (*L. lactis*). Positive clones identified by clearing zones or fluorescence were re-assessed to confirm activities and inserts sequenced (accession numbers see Supporting Information Table S1). Bioinformatic sequence analyses and databases used are detailed in supplemental methods.

Cloning of GH9 encoding genes from *Coprococcus*-related species

Cloning procedures are given in detail in supplemental methods. Briefly, four genes encoding putative GH9 enzymes from *C. eutactus* ART55/1 and *Coprococcus* sp. L2-50 were cloned into shuttle plasmid pTRKL2 and transformed into *E. coli* XL1 Blue and *L. lactis* MG1363. For *Coprococcus* sp. L2-50 gene GH9/S (WP_008400439), synthetic gene constructs containing different domains (Supporting Information Fig. S5) were overexpressed in *E. coli* BL21(DE3) and enzyme activities were determined as described below.

Enzyme activity assays

For GH9 constructs cloned into *E. coli* XL1 Blue and *L. lactis* MG1363, freshly grown overnight culture (10 µl) was pipetted onto the surface of lichenan-, CMC- and xylan-containing agar plates, allowed to incubate overnight and stained with Congo Red. In order to increase the contrast, the plates were flooded with 50 mM acetic acid to turn the background towards the blue colour instead of pale orange.

For determination of enzyme activities in liquid culture and cellular localisation in *E. coli* XL1 Blue and *L. lactis* MG1363, freshly grown overnight cultures of recombinant clones were analysed. The enzyme activity was determined by reducing sugar assay following the Lever method (Lever, 1977), using supernatants and cell-free extracts prepared from three independently grown cultures (for details see supplemental methods). Enzyme activity was determined by measuring the amount of reducing sugar released by the fractions incubated with CMC, lichenan or β-glucan (each at 0.5% w/v) as substrates at 37°C.

For enzyme activities in *E. coli* BL21 (DE3) containing recombinant genes, *C. eutactus* ART55/1 and *Coprococcus* sp. L2-50, cells were harvested by centrifugation, washed in 20 ml and resuspended in 5 ml of sodium phosphate buffer (50 mM, pH 6.5). Cell extracts were prepared by bead-beating and enzyme activities determined by Lever assay (Lever, 1977). The details are provided in supplemental methods.

Gene expression and proteomic analysis of *C. eutactus* ART55/1

Full details of gene expression and proteomics methods are provided in supplemental methods. Briefly, triplicate cultures of *C. eutactus* ART55/1 grown on either glucose, cellobiose, β-glucan, lichenan or glucomannan were harvested during exponential phase (optical density 0.45–0.84, Supporting Information Table S4) and RNA and protein fractions were prepared.

For RNA sequencing, libraries were prepared after ribosomal RNA depletion and sequenced using the High Output 1X75 kit on the Illumina NextSeq 500 platform with v2 chemistry, producing 75 bp single end reads. In total, between 22 118 233 and 40 885 793 reads were produced per sample after quality filtering (99.9% of the raw read on average, Supporting Information Table S4). Raw sequencing data have been deposited in the Array Express database under the E-MTAB-8048. Reads were aligned against the reference genome for *C. eutactus* ART55/1 (FP929039.1, between 88.78% and 90.40% of the filtered reads, Supporting Information Table S4) and counted at gene locations (57.52%–60.00% counted,

Supporting Information Table S4). For differential gene expression analysis, genes that had a CPM (count per million) value of more than one in three or more samples were kept for analysis, and all other genes were removed as low count genes, leaving 2022 genes for analysis. Differential expression analysis was performed using a generalized linear model with contrasts made between glucose (as the baseline) and all other carbon sources and setting significance at false discovery rate (FDR) < 0.05 and Log fold change (LogFC) > 1.

For functional analysis, significant differentially expressed gene sequences were isolated from the genome assembly and compared to the NCBI non-redundant protein database and to the InterPro protein signature database. The results of these searches were analysed with Blast2GO (version 5.2.5) (Conesa *et al.*, 2005) where gene ontology (GO) terms were assigned to 1669 genes. GO enrichment analysis was carried out with Blast2GO using a Fisher Exact Test.

Protein digestion was carried out with porcine trypsin. Peptides were desalted and analysed by LC–MS as previously described (Herrero-de-Dios *et al.*, 2018) using a Q Exactive Plus/Ultimate 3000RSLC nanoLC-MS system (Thermo Fisher Scientific, Hemel Hempstead, UK) to which a 25 cm long PepMap RSLC C18 nano column (internal diameter 75 µm) was fitted. Peak identification and quantification was carried out using MaxQuant (version 1.6.3.4) (Cox & Mann, 2008) with comparisons made to *C. eutactus* ART55/1 reference protein sequences downloaded from NCBI (https://www.ncbi.nlm.nih.gov/genome/13745?genome_assembly_id=175605). Parameters were set to calculate LFQ, as well as to identify potential contaminant proteins from media. In total, 919 proteins were identified by MaxQuant. Nineteen potential contaminants and 10 reverse sequence control proteins were removed before further analysis, leaving 891 proteins. LFQ values were log2 converted and then analysed for differential expression using LIMMA (version 3.38.3) (Ritchie *et al.*, 2015) with an identical linear model as to that used with the RNA sequencing analysis, treating glucose as the baseline for comparisons.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol *et al.*, 2019) partner repository with the dataset identifier PXD014174.

Acknowledgements

Freda M. Farquharson, Harry J. Flint and Petra Louis received financial support from the Scottish Government Rural and Environment Science and Analytical Services Division (RESAS). Anna M. Alessi received a PhD scholarship jointly funded by the Institute of Food Research (Norwich, UK, now Quadram Institute) and the Rowett Institute (University of

Aberdeen, Aberdeen, UK). The authors wish to thank Gill Campbell and Pauline Young for their assistance with colony picking and screening metagenomic libraries, Brennan Martin at the Centre for Genome Enabled Biology and Medicine, University of Aberdeen, for performing the RNA sequencing, Evelyn Argo and Craig Pattinson for advice on sample preparation for proteomics work, Bekhal Kareem Sharif and Manon Le Merrer for help with growth experiments, Sylvia Duncan for the gift of acid-swollen cellulose and for providing strains *C. comes* A2-232 and SL7/1, Pat Bain for help with figure formats and Mike Peck and Bruce Pearson at the Institute for Food Research Norwich for helpful discussions.

References

- Aspeborg, H., Coutinho, P.M., Wang, Y., Brumer, H., and Henrissat, B. (2012) Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evol Biol* **12**: 186.
- Barcenilla, A. (1999) Diversity of the butyrate-producing microflora on the human gut. PhD Thesis. Aberdeen, UK: Robert Gordon University.
- Barcenilla, A., Pryde, S.E., Martin, J.C., Duncan, S.H., Stewart, C.S., Henderson, C., and Flint, H.J. (2000) Phylogenetic relationships of butyrate-producing bacteria from the human gut. *Appl Environ Microbiol* **66**: 1654–1661.
- Burdette, L.A., Leach, S.A., Wong, H.T., and Tullman-Ercek, D. (2018) Developing gram-negative bacteria for the secretion of heterologous proteins. *Microb Cell Fact* **17**: 196.
- Cann, I., Bernardi, R.C., and Mackie, R.I. (2016) Cellulose degradation in the human gut: *Ruminococcus champanellensis* expands the cellulose paradigm. *Environ Microbiol* **18**: 307–310.
- CAZyedia Consortium. (2018) Ten years of CAZyedia: a living encyclopedia of carbohydrate-active enzymes. *Glycobiology* **28**: 3–8.
- Cecchini, D.A., Laville, E., Laguerre, S., Robe, P., Leclerc, M., Dore, J., et al. (2013) Functional metagenomics reveals novel pathways of prebiotic breakdown by human gut bacteria. *PLoS ONE* **8**: e72766.
- Chassard, C., Delmas, E., Robert, C., and Bernalier-Donadille, A. (2010) The cellulose-degrading microbial community of the human gut varies according to the presence or absence of methanogens. *FEMS Microbiol Ecol* **74**: 205–213.
- Chassard, C., Delmas, E., Robert, C., Lawson, P.A., and Bernalier-Donadille, A. (2011) *Ruminococcus champanellensis* sp. nov., a cellulose-degrading bacterium from the human gut microbiota. *Int J Syst Evol Microbiol* **62**: 138–143.
- Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., and Robles, M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**: 3674–3676.
- Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**: 1367–1372.
- Devillard, E., Goodheart, D.B., Karnati, S.K.R., Bayer, E.A., Lamed, R., Miron, J., et al. (2004) *Ruminococcus albus* 8 mutants defective in cellulose degradation are deficient in two processive endocellulases, Cel48A and Cel9B, both of which possess a novel modular architecture. *J Bacteriol* **186**: 136–145.
- Dobrijevic, D., di Liberto, G., Tanaka, K., de Wouters, T., Dervyn, R., Boudebouze, S., et al. (2013) High-throughput system for the presentation of secreted and surface-exposed proteins from gram-positive bacteria in functional metagenomics studies. *PLoS one* **8**: e65956.
- Duncan, S.H., Barcenilla, A., Stewart, C.S., Pryde, S.E., and Flint, H.J. (2002) Acetate utilization and butyryl coenzyme a (CoA):acetate-CoA transferase in butyrate-producing bacteria from the human large intestine. *Appl Environ Microbiol* **68**: 5186–5190.
- Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**: 783–791.
- Flint, H.J., Bayer, E.A., Rincon, M.T., Lamed, R., and White, B.A. (2008) Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nat Rev Microbiol* **6**: 121–131.
- Flint, H.J., Scott, K.P., Duncan, S.H., Louis, P., and Forano, E. (2012a) Microbial degradation of complex carbohydrates in the gut. *Gut Microbes* **3**: 289–306.
- Flint, H.J., Scott, K.P., Louis, P., and Duncan, S.H. (2012b) The role of the gut microbiota in nutrition and health. *Nat Rev Gastroenterol Hepatol* **9**: 577–589.
- Freudl, R. (2018) Signal peptides for recombinant protein secretion in bacterial expression systems. *Microb Cell Fact* **17**: 52.
- Gabor, E.M., Alkema, W.B.L., and Janssen, D.B. (2004) Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environ Microbiol* **6**: 879–886.
- Gagic, D., Ciric, M., Wen, W.X., Ng, F., and Rakonjac, J. (2016) Exploring the secretomes of microbes and microbial communities using filamentous phage display. *Front Microbiol* **7**: 429.
- Gasson, M.J. (1983) Plasmid complements of *Streptococcus lactis* NCDO 712 and other lactic streptococci after protoplast-induced curing. *J Bacteriol* **154**: 1–9.
- Handelsman, J. (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* **68**: 669–685.
- Herrero-de-Dios, C., Day, A., Tillmann, A.T., Kastora, S.L., Stead, D., Salgado, P.S., et al. (2018) Redox regulation, rather than stress-induced phosphorylation, of a Hog1 mitogen-activated protein kinase modulates its nitrosative-stress-specific outputs. *mBio* **9**: e02229-17.
- Holdeman, L.V., and Moore, W.E.C. (1974) New genus, *Coprococcus*, twelve new species, and emended descriptions of four previously described species of bacteria from human feces. *Int J Syst Bacteriol* **24**: 260–277.
- Howlett, J.F., Betteridge, V.A., Champ, M., Craig, S.A.S., Meheust, A., and Jones, J.M. (2010) The definition of dietary fiber – discussions at the ninth Vahouny fiber symposium: building scientific agreement. *Food Nutr Res* **54**: 5750.
- Kumar, S., Stecher, G., Li, M., Nnyaz, C., and Tamura, K. (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* **35**: 1547–1549.

- Kurokawa, J., Hemjinda, E., Arai, T., Kimura, T., Sakka, K., and Ohmiya, K. (2002) *Clostridium thermocellum* cellulase CelT, a family 9 endoglucanase without an Ig-like domain or family 3c carbohydrate-binding module. *Appl Microbiol Biotechnol* **59**: 455–461.
- Lever, M. (1977) Carbohydrate determination with 4-hydroxybenzoic acid hydrazide (PAHBAH): effect of bis-muth on the reaction. *Anal Biochem* **81**: 21–27.
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucl Acids Res* **42**: D490–D495.
- Louis, P., Duncan, S.H., McCrae, S.I., Millar, J., Jackson, M. S., and Flint, H.J. (2004) Restricted distribution of the butyrate kinase pathway among butyrate-producing bacteria from the human colon. *J Bacteriol* **186**: 2099–2106.
- Louis, P., and Flint, H.J. (2017) Formation of propionate and butyrate by the human colonic microbiota. *Environ Microbiol* **19**: 29–41.
- McMahon, M.D., Guan, C., Handelsman, J., and Thomas, M.G. (2012) Metagenomic analysis of *Streptomyces lividans* reveals host-dependent functional expression. *Appl Environ Microbiol* **78**: 3622–3629.
- Mingardon, F., Chanal, A., Tardif, C., and Fierobe, H.P. (2011) The issue of secretion on heterologous expression of *Clostridium cellulolyticum* cellulase-encoding genes in *Clostridium acetobutylicum* ATCC 824. *Appl Environ Microbiol* **77**: 2831–2838.
- Miyazaki, K., Martin, J.C., Marinsek-Logar, R., and Flint, H.J. (1997) Degradation and utilization of xylans by the rumen anaerobe *Prevotella bryantii* (formerly *P. ruminicola* subsp. *bovis*) B14. *Anaerobe* **3**: 373–381.
- Moracci, M. (2019) Glycoside Hydrolase Family 42 in CAZyedia. URL <http://www.cazypedia.org/>.
- Morais, S., Ben David, Y., Bensoussan, L., Duncan, S.H., Koropatkin, N.M., Martens, E.C., et al. (2016) Enzymatic profiling of cellulosomal enzymes from the human gut bacterium, *Ruminococcus champanellensis*, reveals a fine-tuned system for cohesin-dockerin recognition. *Environ Microbiol* **18**: 542–556.
- Nylund, L., Nermes, M., Isolauri, E., Salminen, S., de Vos, W.M., and Satokari, R. (2015) Severity of atopic disease inversely correlates with intestinal microbiota diversity and butyrate-producing bacteria. *Allergy* **70**: 241–244.
- O'Sullivan, D.J., and Klaenhammer, T.R. (1993) High- and low-copy-number *Lactococcus* shuttle cloning vectors with features for clone screening. *Gene* **137**: 227–231.
- Pereira, J.H., Saprà, R., Volponi, J.V., Kozina, C.L., Simmons, B., and Adams, P.D. (2009) Structure of endoglucanase Cel9A from the thermoacidophilic *Alicyclobacillus acidocaldarius*. *Acta Crystallogr D Biol Crystallogr* **65**: 744–750.
- Pereira, J.H., Chen, Z., McAndrew, R.P., Saprà, R., Chhabra, S.R., Sale, K.L., et al. (2010) Biochemical characterization and crystal structure of endoglucanase Cel5A from the hyperthermophilic *Thermotoga maritima*. *J Struct Biol* **172**: 372–379.
- Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J., et al. (2019) The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* **47**: D442–D450.
- Pontes, D.S., De Azevedo, M.S.P., Chatel, J.-M., Langella, P., Azevedo, V., and Miyoshi, A. (2011) *Lactococcus lactis* as a live vector: heterologous protein production and DNA delivery systems. *Protein Exp Purif* **79**: 165–175.
- Reichardt, N., Duncan, S.H., Young, P., Belenguer, A., McWilliam Leitch, C., Scott, K.P., et al. (2014) Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. *ISME J* **8**: 1323–1335.
- Reichardt, N., Vollmer, M., Holtrop, G., Farquharson, F., Wefers, D., Bunzel, M., et al. (2018) Specific substrate-driven changes in human faecal microbiota composition contrast with functional redundancy in short-chain fatty acid production. *ISME J* **12**: 610–622.
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015) Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucl Acids Res* **43**: e47.
- Robert, C., Chassard, C., Lawson, P.A., and Bernalier-Donadille, A. (2007) *Bacteroides cellulolyticus* sp. nov., a cellulolytic bacterium from the human gut microbial community. *Int J Syst Evol Microbiol* **57**: 1516–1520.
- Saitou, N., and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406–425.
- Sheridan, P.O., Martin, J.C., Lawley, T.D., Browne, H.P., Harris, H.M.B., Bernalier-Donadille, A., et al. (2016) Polysaccharide utilization loci and nutritional specialization in a dominant group of butyrate-producing human colonic firmicutes. *Microb Genom* **2**: 1–16.
- Slavin, J. (2013) Fiber and prebiotics: mechanisms and health benefits. *Nutrients* **5**: 1417–1435.
- Sun, Y., Su, Y., and Zhu, W. (2016) Microbiome-metabolome responses in the cecum and colon of pig to a high resistant starch diet. *Front Microbiol* **7**: 779.
- Tamura, K., Nei, M., and Kumar, S. (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci U S A* **101**: 11030–11035.
- 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.
- Tuncil, Y.E., Xiao, Y., Porter, N.T., Reuhs, B.L., Martens, E. C., and Hamaker, B.R. (2017) Reciprocal prioritization to dietary glycans by gut bacteria in a competitive environment promotes stable coexistence. *mBio* **8**: e01068-17.
- Uchiyama, T., and Miyazaki, K. (2009) Functional metagenomics for enzyme discovery: challenges to efficient screening. *Curr Opin Biotechnol* **20**: 616–622.
- Valles-Colomer, M., Falony, G., Darzi, Y., Tigchelaar, E.F., Wang, J., Tito, R.Y., et al. (2019) The neuroactive potential of the human gut microbiota in quality of life and depression. *Nat Microbiol* **4**: 623–632.
- Vogel, C., and Marcotte, E.M. (2012) Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* **13**: 227–232.
- Wilson, D. B., and Urbanowicz, B. (2019) Glycoside Hydrolase Family 9 in CAZyedia. URL <http://www.cazypedia.org/>.

- Ze, X., Ben David, Y., Laverde-Gomez, J., Dassa, B., Sheridan, P.O., Duncan, S.H., *et al.* (2015) Unique organization of extracellular amylases into 'amylosomes' in the resistant starch-utilizing human colonic Firmicutes bacterium *Ruminococcus bromii*. *mBio* **6**: e01425-15.
- Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., *et al.* (2018) dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucl Acids Res* **46**: W95–W101.
- Zhang, S., Zubay, G., and Goldman, E. (1991) Low-usage codons in *Escherichia coli*, yeast, fruit fly and primates. *Gene* **105**: 61–72.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1 Sequence analysis of clones recovered by a functional screening of human gut microbiome metagenomic libraries.

Table S2. Enzyme activity of GH9 enzymes from *C. eutactus* ART55/1 and *Coprococcus* sp. L2-50 expressed in *E. coli* XL1 Blue and *L. lactis* MG1363 supernatant and cell-free extract.

Table S3. Reducing sugar assay of cell extracts and supernatants of cultures of *C. eutactus* ART55/1 and *Coprococcus* sp. L2-50 grown on glucose, cellobiose or β -glucan to early stationary phase.

Table S4: Individual culture and sample statistics of gene expression and proteomics analysis.

Table S5. Log fold change (logFC) and adjusted significance values (FDR and adj.P.Val) for RNA sequencing and proteomics analyses of *C. eutactus* ART55/1 genes.

Table S6. Significantly enriched Gene Ontology (GO) terms identified by Blast2GO (FDR < 0.05) *C. eutactus* ART55/1 genes.

Table S7. CAZyme of *C. eutactus* ART55/1 and *Coprococcus* sp. L2-50.

Table S8. Primers used for amplification of GH9 genes from *Coprococcus*-related species.

Fig. S1 Phylogenetic analysis of the family 9 catalytic modules of *C. eutactus* ART55/1 and *Coprococcus*

sp. L2-50 GH9 genes (highlighted by blue and purple dots, respectively). The phylogenetic trees show the relationship of the catalytic domains of ART_GH9/S and L2-50_GH9/S proteins (A) and ART_GH9/L and L2-50_GH9/L (B) proteins. The amino acid sequences of the catalytic domains were retrieved from NCBI following BlastP analysis. The evolutionary history was inferred using the Neighbour-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method and are in the units of the number of amino acid differences per site. The analysis involved 68 (A) and 44 (B) amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 376 (A) and 348 (B) positions in the final dataset. Evolutionary analyses were conducted in MEGA7.

Fig. S2. Carbohydrate utilization of *C. eutactus* ART55/1, *Coprococcus* sp. L2-50, *C. eutactus* ATCC 27759, *C. comes* ATCC 27758, *C. comes* A2-232, *C. comes* SL7/1 and *C. catus* GD7, on different substrates.

A. Medium pH drop after growth in 96-well plates after 48 h of incubation (mean and standard deviation of triplicate cultures).

B. Growth curves during growth in 96-well plates. *C. catus* GD/7 does not grow well on glucose, but the preculture showed an optical density increase during the hour before inoculation and had reached OD 0.25 at inoculation of the growth experiment.

Fig. S3. Relationship of all differentially expressed *C. eutactus* ART55/1 genes (DEGs, 299 genes in total) between different growth substrates, expressed as logFC. Other comparisons are shown in Fig. 5C.

Fig. S4. Venn diagrams comparing significant differential expression of *C. eutactus* ART55/1 genes and proteins in glucose vs. β -glucan (A), glucose vs. cellobiose (B), glucose vs. glucomannan (C), and glucose vs. lichenan (D).

Fig. S5. Cloning strategy for generation of synthetic gene constructs of *Coprococcus* sp. L2-50 GH9/S.

Appendix S1: Supplemental methods