1 **Title**

2 Butyrate producing Clostridiales utilize distinct human milk oligosaccharides correlating to early
3 colonization and prevalence in the human gut

4 **Authors**

5 Michael Jakob Pichler[1#], Chihaya Yamada[2#], Bashar Shuoker[1,7], Maria Camila Alvarez-Silva[6],
6 Aina Gotoh[3], Maria Louise Leth[1], Erwin Schoof[1], Toshihiko Katoh[3], Mikiyasu Sakanaka[4],
7 Takane Katayama[3,4], Chunsheng Jin[5], Niclas G. Karlsson[5], Manimozhiyan Arumugam[6,]
8 Shinya Fushinobu[2], Maher Abou Hachem[1,*]

9 Affiliations

10 [1]Department of Biotechnology and Biomedicine, Technical University of Denmark, Lyngby,
11 2800, Denmark
12 [2]Department of Biotechnology, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo, 113-
13 8657, Japan
14 [3]Graduate School of Biostudies, Kyoto University, Sakyo-ku, Kyoto, 606-8502, Japan
15 [4]Faculty of Bioresources and Environmental Sciences, Ishikawa Prefectural University,
16 Nonoichi, Ishikawa, 921-8836, Japan
17 [5]Department of Medical Biochemistry, University of Gothenburg, Gothenburg, 413 90,Sweden
18 [6]The Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and
19 Medical Sciences, University of Copenhagen, Copenhagen,2200, Denmark
20 [7]Division of Biotechnology, Department of Chemistry, Lund University, Lund, 22100, Sweden
21 *Correspondence: maha@bio.dtu.dk
22 #These authors contributed equally to this work

23
24
25 Contributions

26 Bacterial growth studies were performed by M.J.P. Proteomic analyses were done by M.J.P and
27 E.S. Protein characterization was done by M.J.P and M.L.L. Enzymatic characterization of
28 *Ri*Le$^{a/b}$136 was performed by A.G.,To.K., M.S. and Ta.K. Mucin preparation was performed by
29 B.S. and M.J.P. Mucin glycomics were performed by B.S., C.J. and N.G.K. Protein X-ray
30 crystallography was performed by C.Y. and S.F. Metagenome analysis were performed by C.A-
31 S and M.A. Experiments were designed by M.J.P and M.A.H. The manuscript was drafted by
32 M.J.P and M.A.H and finalized with contributions of all authors.

33

34

35

Abstract

37 The early life human gut microbiota exerts life-long health effects on the host, but the mechanisms
38 underpinning its assembly remain elusive. Particularly, the early colonization of Clostridiales from
39 the *Roseburia-Eubacterium* group, associated with protection from colorectal cancer, immune-
40 and metabolic disorders is enigmatic. Here we unveil the growth of *Roseburia* and *Eubacterium*
41 members on human milk oligosaccharides (HMOs) using an unprecedented catabolic apparatus.
42 The described HMO pathways and additional glycan utilization loci confer co-growth with
43 *Akkermansia muciniphilia* via cross-feeding and access to mucin *O*-glycans. Strikingly, both, HMO
44 and xylooligosaccharide pathways, were active simultaneously attesting an adaptation to a mixed
45 HMO-solid food diet. Analyses of 4599 *Roseburia* genomes underscored the preponderance of
46 HMO pathways and highlighted different HMO utilization phylotypes. Our revelations provide a
47 possible rationale for the early establishment and resilience of butyrate producing Clostridiales
48 and expand the role of specific HMOs in the assembly of the early life microbiota.

49

Introduction

51 The human gut microbiota (HGM) is a key determinant of health[1–3]. Orthogonal transfer from the
52 mother contributes markedly to the establishment of this community shortly after birth[4,5]. The HGM
53 develops dynamically during infancy until a resilient adult-like community is formed after 2-3 years
54 of life[6–8]. The early life microbiota plays a role in the maturation of the host's endocrine, metabolic
55 and immune system[9], and the composition of this consortium appears to be associated with life-
56 long health effects[10–12]. Therefore, understanding the factors that define the HGM structure during
57 infancy is critical for minimizing the risk for a range of metabolic, inflammatory and
58 neurodegenerative disorders, all associated to specific HGM signatures[13,14].

59 Dietary glycans resistant to digestion by human enzymes are a major driver that shapes the
60 developing HGM[6,15]. This is emphasized by the dominance of *Bifidobacterium* in breast-fed
61 infants[7,8], attributed to the competitiveness of distinct members of this genus in the utilization of
62 human milk oligosaccharides (HMOs)[16,17]. Indeed, the most prominent changes in the infant
63 microbiota occur during weaning and the introduction of solid food[6,7], whereby bifidobacteria are
64 replaced by Firmictues as the top abundant phylum of the mature HGM. This compositional shift
65 is accompanied by a notable longitudinal increases in the concentrations of the short chain fatty
66 acids (SCFAs) propionate and butyrate (generated from carbohydrate fermentation) during and
67 after weaning[18].

68 Butyrate exerts immune-modulatory activities[19] and is associated with a lowered risk of colon
69 cancer, atherosclerosis, and enteric colitis[20,21]. The bacterial production of butyrate is largely
70 ascribed to Firmicutes *Clostridium* cluster IV and *Clostridium* cluster XIVa that includes members
71 of the *Roseburia-Eubacterium* group (Lachnospiraceae family, Clostridiales order), which are
72 abundant and prevalent members of the adult HGM[22,23]. By contrast, the abundance of *Roseburia*
73 spp. is decreased in patients suffering from metabolic, inflammatory and cardiovascular
74 diseases[24–27]. Although butyrate producers are established by the first year of life[27], the
75 mechanisms underpinning their early appearance (and prevalence) remain unknown.

76 The evolution of uptake and enzymatic systems that support competitive growth of
77 *Bifidobacterium* species on HMOs[17] reflects a successful adaptation to the intestines of breast fed
78 infants. We hypothesized that other taxonomic groups, which possess metabolic capabilities that
79 target HMOs, may have an early advantage in the colonization of the infant gut during infancy.

80 The early emergence of *Roseburia-Eubacterium* members that comprise the main group of
81 butyrate producing bacteria in the human adult gut offers a suitable model group to evaluate this
82 hypothesis. Genomic analyses were suggestive of the presence of putative HMOs utilization gene
83 clusters in *Roseburia* and *Eubacterium* strains. Growth on distinct HMOs (pure or in a complex
84 mixture from mothers milk) and differential proteomics from HMO growth experiments provided
85 compelling evidence for the molecular basis of HMO utilization by this taxonomic group. We
86 corroborated these finding by molecular characterization of the enzymes and transport proteins
87 that confer growth on HMOs. These analyses disclosed an unprecedented enzymatic activity and
88 a previously unknown structural fold highlighting the uniqueness in the enzymology of HMOs
89 utilization by this taxonomic group. We also showed that the unveiled catabolic pathways support
90 cross-feeding on mucin *O*-glycans in co-culture with the model mucin degrader *Akkermansia*
91 *muciniphila*. Analyses of the metagenome of *Roseburia*, showed a striking conservation and
92 preponderance of the HMO utilization pathways across the genus, underscoring their importance
93 for adaptation to the human gut. This study provides unprecedented mechanistic details into
94 pathways that may contribute to the early colonization and the resilience of key butyrate producing
95 Clostridiales by mediating the catabolism of distinct HMOs and host *O*-glycans.

96

Results

Roseburia hominis and _Roseburia inulinivorans_ possess conserved gene loci that support growth on distinct human milk oligosaccharides (HMOs)

We hypothesized that HMO utilization may confer an early advantage in the assembly of early life HGM. Genomic analyses of butyrate producers from Lachnospiraceae identified distant homologs of the recently discovered glycoside hydrolase family 136 (GH136) in the Carbohydrate Active enZyme database (www.cazy.org) (Supplementary Fig. 1). This family was assigned based on the lacto-_N_-biosidase LnbX from _Bifidobacterium longum_ subsp. _longum_ JCM 1217[28], which cleaves the key HMO lacto-_N_-tetraose (LNT) to lacto-_N_-biose (LNB) and lactose (EC 3.2.1.140; Supplementary Table 1).

We selected two _Roseburia_ strains and one from _Eubacterium_, all having GH136-like genes, to examine their HMO utilization capabilities.

Significant growth was observed for _Roseburia hominis_ DSM 16839 ($p<4.0 \times 10^{-4}$) and _Roseburia inulinivorans_ DSM 16841 ($p<1.3 \times 10^{-4}$) after 24 h on media supplemented with HMOs from mother milk, but the growth of _R. inulinivorans_ was more efficient (μ_{max}=0.30 ± 0.01 h^{-1}). Next, we carried out growth on building blocks present in HMOs and related oligomers from _O_-glycoconjugates (Fig. 1a-d). _R. hominis_ grew efficiently on LNT (μ_{max}=0.22 ± 0.02 h^{-1}), its LNB unit (μ_{max}=0.16 ± 0.01 h^{-1}) and the mucin derived galacto-_N_-biose (GNB) (μ_{max}=0.21 ± 0.02 h^{-1}). By contrast, _R. inulinivo_rans grew better on LNB and GNB relative to LNT. Growth on LNT was also shared by the taxonomically related _Eubacterium ramulus_ DSM 15684 from Eubacteriaceae. _R. inulinivorans_ was distinguished by growth on sialic acid (Neu5Ac), abundant in HMOs and glycoconjugates (Fig. 1d).

To unravel the basis of growth on HMOs, we analyzed the differential proteomes of _R. hominis_ and _R. inulinivorans_ on LNT and the HMO mixture, respectively, relative to glucose. For _R. hominis_ and _R. inulinivorans_, 15 and 62 proteins, respectively, were significantly upregulated (log$_2$ fold change > 2). These differential proteomes were dominated by carbohydrate metabolism proteins, especially products of two loci, both encoding an ATP-binding cassette (ABC) transporter, GH112 and GH136 enzymes with putative HMO activities, as well as sensory and transcriptional regulators (Fig. 1f,g). The HMO locus of _R. inulinivorans_ is extended with two fucosidases of GH29 and GH95. The specificity-determining solute binding proteins (SBPs) of the ABC transporters of _R. hominis_ (_Rh_LNBBP) and _R. inulinivorans_ (_Ri_Le$^{a/b}$BP) were the first and fifth top-upregulated proteins in the HMO proteomes, respectively. In addition, the GH112 LNB/GNB phosphorylases were within the top 3 and 12 upregulated proteins in _R. hominis_ and _R. inulinivorans_, respectively. In _R. inulinivorans_ two additional loci encoding sialic acid and fucose catabolism proteins, were also upregulated (Supplementary Fig. 3).
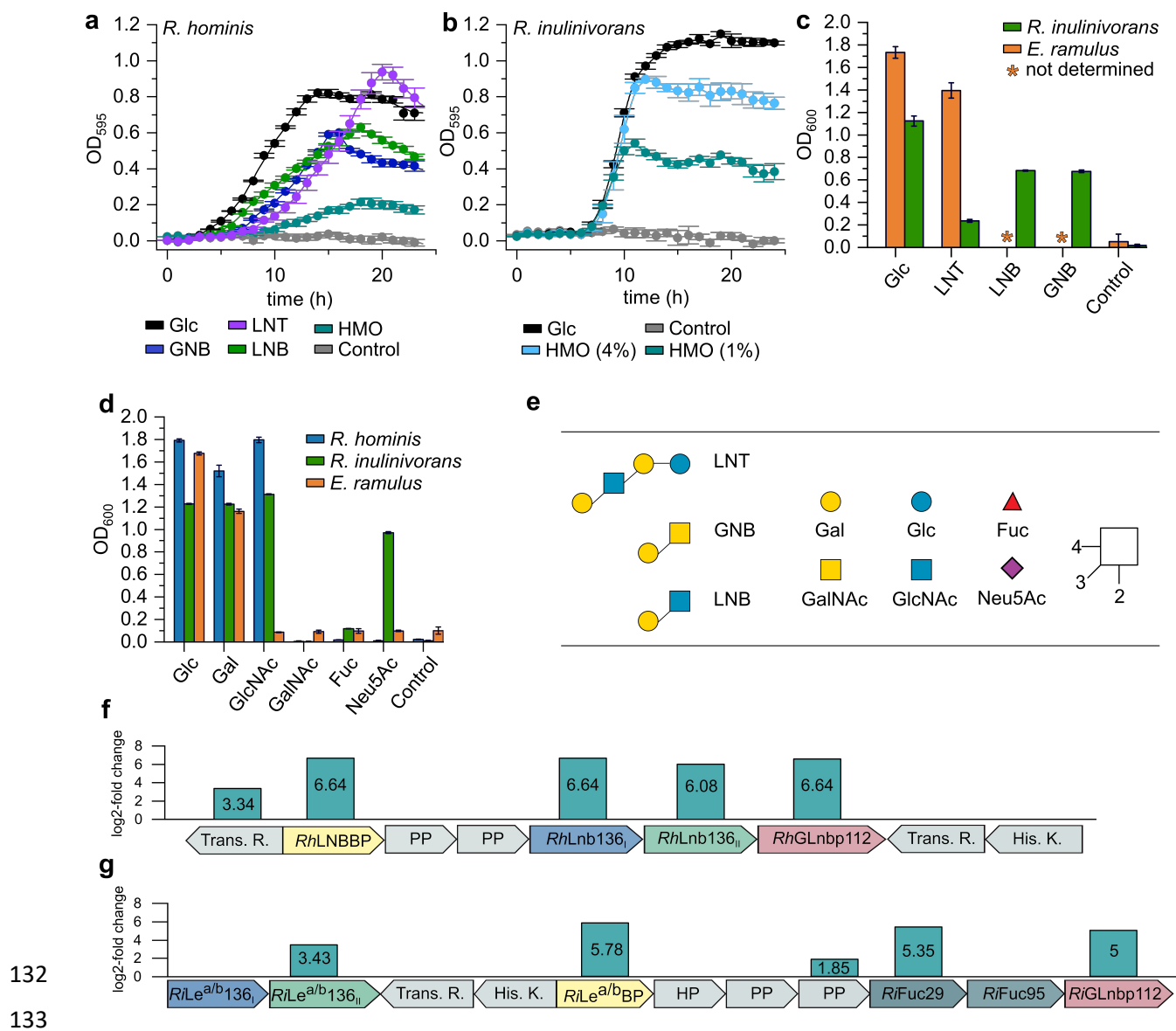
Fig. 1: Growth of *R. hominis*, *R. inulinivorans* and *E. ramulus* on HMOs and upregulation of core HMOs utilization loci: Growth curves of *R. hominis* (**a**) and *R. inulinivorans* (**b**) on glucose, LNT, GNB, LNB, and/or purified HMOs from mothers milk compared to a no-carbon source controls over 24 h. **c**, Growth levels of *R. inulinivorans* on LNT, LNB, GNB and of *E. ramulus* on LNT within 24 h including glucose and a non-carbon source controls. **d**, Growth of *R. hominis*, *R. inulinivorans* and *E. ramulus* on monosaccharides from HMOs and mucin after 24 h. The growth analyses (**a-d**) on media supplemented with 0.5 % (w/v) carbohydrates (except for *R. inulinivorans* on 1% (w/v) and 4% (w/v) purified HMOs from mothers milk) are means of triplicates with standard deviations. **e**) HMO and mucin-derived oligo- and monosaccharides used for the growth analyses in (**a-d**). The core HMO utilization loci in *R. hominis* (**f**) and *R. inulinivorans* (**g**) identified form proteomic analyses of cells growing on LNT and HMOs from mothers milk, respectively, relative to glucose. The proteomic analyses (**f-g**) were performed in biological triplicates and the log$_2$-fold change from the label free quantification of upregulated gene products is shown. Glycan structures presentation according to Symbol Nomenclature for Glycans (SNFG) (https://www.ncbi.nlm.nih.gov/glycans/snfg.html).
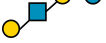
Transport proteins of _R. hominis_ and _R. inulinivorans_ capture HMO blocks and host-derived oligosaccharides

The proteomic analyses highlighted the putative protein apparatus required for growth on HMOs. The solute binding proteins (SBPs) of two ABC transporters in _R. hominis_ and _R. inulinivorans_ were within the top 8% upregulated proteins, hinting their involvement in uptake of HMOs. Both SBPs recognized distinct HMOs and ligands from host-glycans (Table 1, Supplementary Tables 2 and 3, Supplementary Fig. 4). The _R. hominis_ SBP (LNB-binding protein, _Rh_LNBBP) displays a preference for LNB followed by about 3.5 fold lower affinities towards GNB and LNT. By contrast, the SBP of _R. inulinivorans_ (Le[a/b] binding protein, _Ri_Le[a/b]BP) prefers fucosyl-decorated Lewis b (Le[b]) tetraose and Lewis a (Le[a]) triose followed by LNB and GNB, whereas no binding was detected to LNT (Table 1). The loss of the fucosyl unit at the terminal reducing GlcNAc reduced the affinity of _Ri_Le[a/b]BP about 5-fold for blood group H antigen triose type I (H triose type I) relative to Le[b] tetraose. _Ri_Le[a/b]BP had no affinity for lacto-_N_-neotetroase (LN_n_T) and blood group A antigen triose (A triose). Lactose and 2'-fucosyllactose (2'-FL) were not recognized by either SBP. These results established the capture of specific HMOs and related ligands by these SBPs and the differentiation of their specificities, e.g. preference of _Ri_Le[a/b]BP to fucosylated ligands at the terminal reducing GlcNAc.

Table 1: Binding analysis of HMOs and related host-derived oligosaccharides to transport proteins from _R. hominis_ and _R. inulinivorans_.

Ligand	_Rh_LNBBP $K_D{}^a$ (µM)	_Ri_Le[a/b]BP $K_D{}^b$ (µM)	Structure
LNB	2.9 ± 0.3	6.7 ± 0.7	
GNB	11.1 ± 0.1	11 ± 0.9	
Le[b] tetraose	n.d.[c]	1.8 ± 0.1	
Le[a] triose	n.d.	3.2 ± 0.5	
H triose type I	n.d.	11.3 ± 2.5	
LNT	10.3 ± 0.6	n.b.[d]	

[a]K_D was determined by isothermal titration calorimetry (ITC). [b]K_D was determined by surface plasmon resonance (SPR). Both analyses were in duplicates and the K_D values are means with standard deviations. The errors from the fit of a one-binding site model to the binding isotherms are reported for the ITC experiments. [c]n.d.: not determined. [d]n.b.: low affinity precluding reliable determination of the binding constant.

165

166

167

168 Functionally diverse GH136 enzymes confer the initial hydrolysis of HMO blocks and
169 related oligosaccharides within *Roseburia* and *Eubacterium*

170 The difference in transport preferences between *R. hominis* and *R. inulinivorans* was indicative
171 of different routes for the utilization of HMO-blocks and related oligomers.

172 The affinity of *Rh*LNBBP from *R. hominis* for LNT, suggested the uptake and subsequent
173 intracellular degradation. Besides the ABC importer, the HMO utilization locus in *R. hominis*
174 encodes distant homologs of the two proteins reported to be necessary for the heterologous
175 expression of the GH136 lacto-*N*-biosidase from *B. longum*[29]. The homologs *Rh*Lnb136$_I$ (LnbY in
176 *B. longum*) and *Rh*Lnb136$_{II}$ (Fig. 1f and Supplementary Fig. 1) that harbors the catalytic residues
177 (LnbX in *B. longum*) were highly co-upregulated in the LNT proteome of *R. hominis*. Both proteins
178 lacked predicted signal peptides and transmembrane domains (Supplementary Fig. 8a), in
179 contrast to the *B. longum* counterparts, suggesting the intracellular degradation of LNT in *R.*
180 *hominis*. Only co-expression and co-purification of *Rh*Lnb136$_I$ and *Rh*Lnb136$_{II}$ resulted in an
181 active lacto-*N*-biosidase (henceforth *Rh*Lnb136) (Fig. 2b, Supplementary Table 4). These findings
182 and the observed co-upregulation, suggested that a hetero-oligomer of the *Rh*Lnb136$_I$ and
183 *Rh*Lnb136$_{II}$ subunits assembles the catalytically active *Rh*Lnb136. Next, we demonstrated
184 phosphorolysis of LNB and GNB to α-D-galactose-1-phosphate and the corresponding *N*-
185 acetylhexosamines GlcNAc and GalNAc, respectively (Supplementary Fig. 6f), by the GH112
186 GNB/LNB phosphorylase (*Rh*GLnbp112) in the same locus (Fig. 1f and Supplementary Fig. 1).
187 This enzyme has comparable specific activities for LNB and GNB (Supplementary Table 5)
188 consistent with the growth data on these disaccharides. The functional lacto-*N*-biosidase and
189 GNB/LNB phosphorylase further support the HMO catabolism role of the locus.

190 The SBP of *R. inulinivorans* had no measurable affinity for LNT in accord with the poor growth
191 (Fig. 1c). Intriguing differences were also observed in the GH136 homolog from the HMO-
192 upregulated locus in *R. inulinivorans* compared to *Rh*Lnb136: 1) the *Ri*GH136$_I$ subunit has an N-
193 terminal transmembrane domain, 2) a signal peptide was predicted at the N-terminus of
194 *Ri*GH136$_{II}$, 3) the presence of two C-terminal putative carbohydrate binding modules in *Ri*GH136$_{II}$
195 (Supplementary Fig. 8a). Co-expression of *Ri*GH136$_I$ and *Ri*GH136$_{II}$, lacking the transmembrane
196 domain and signal peptide respectively, resulted in an active enzyme with an unprecedented
197 specificity. This enzyme (*Ri*Le$^{a/b}$136) released Lewis a triose or Lewis b tetraose from fucosylated
198 HMOs including lacto-*N*-fucopentaose II (LNFP II), lacto-*N*-difucohexaose I (LNDFH I) and lacto-
199 *N*-difucohexaose II (LNDFH II) (Fig. 2a and Supplementary Fig. 6a). To our knowledge, enzymatic
200 activity on the glycosidic bond at the reducing end of a fucosylated-GlcNAc unit in the above
201 HMOs has not been reported to date. The products of *Ri*Le$^{a/b}$136 are the preferred ligands for
202 *Ri*Le$^{a/b}$BP, suggesting the uptake of these products by the ABC-transporter. Next, we showed that
203 the concerted action of *Ri*Fuc29 and *Ri*Fuc95 that act on α-(1→4) and α-(1→2)-linked L-fucosyl,
204 respectively mediates the complete defucosylation of Le^b tetraose, Le^a triose and H triose type I
205 (Supplementary Fig. 6b-d). Initial defucosylation by *Ri*Fuc29 is required for releasing the 1→2
206 linked L-fucosyl in Le^b tetraose by *Ri*Fuc95. Finally, we showed that the GH112 from *R.*
207 *inulinivorans* (*Ri*GLnbp112) phosphorolyzes LNB and GNB equally efficiently (Supplementary
208 Fig. 6e, Supplementary Table 5).
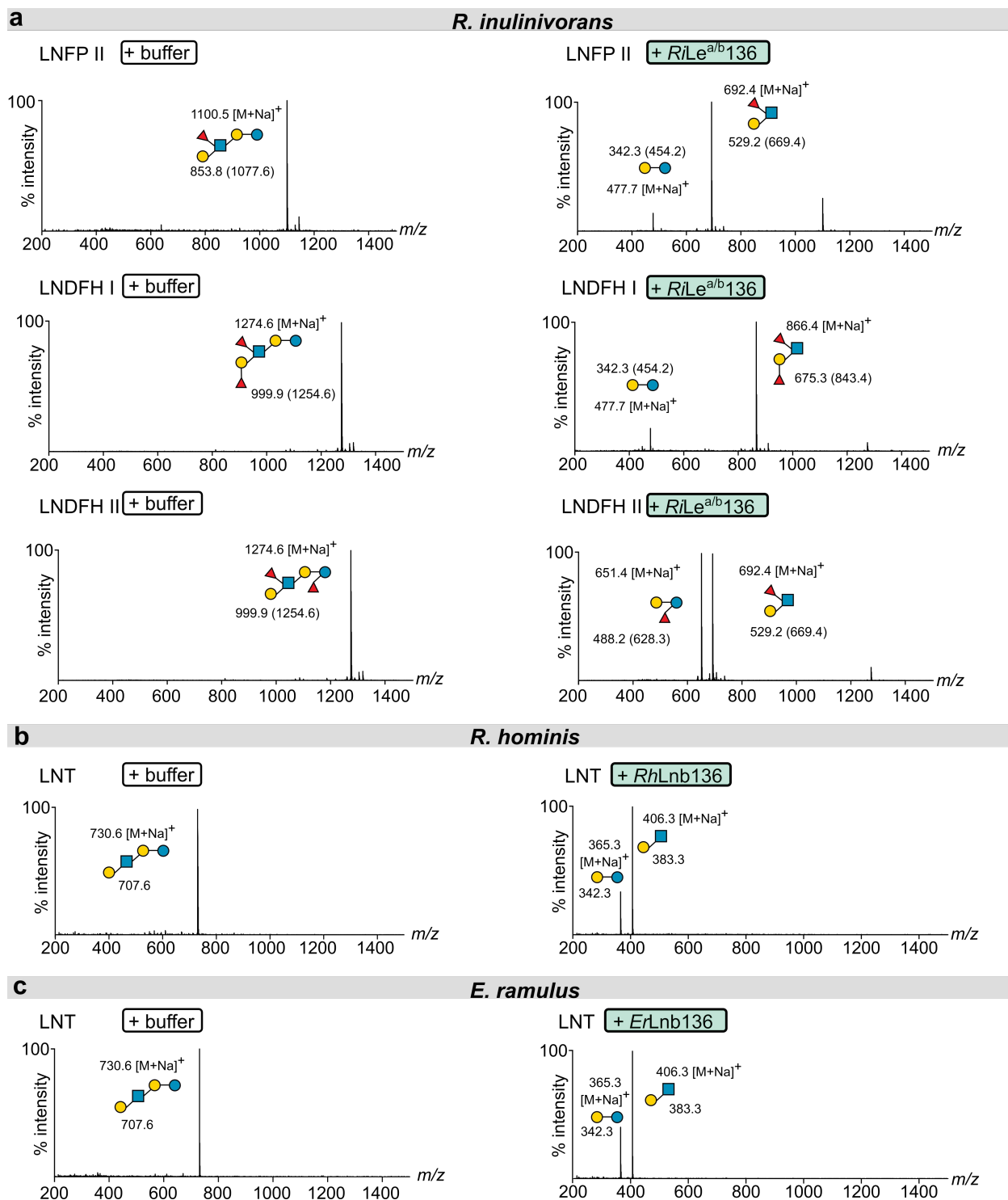
Fig. 2: Different specificities of GH136 enzymes in *Roseburia* and *Eubacterium* mediate the degradation of distinct HMOs. (a), Activity of *Ri*Le$^{a/b}$136 on fucosylated HMOs. **(b)**, Activity of *Rh*Lnb136 on LNT. **(c)**, Activity of *Er*Lnb136 on LNT. **(a-c)**, The hydrolysates were analyzed by MALDI-ToF MS without **(b,c)** or with **(a)** previous permethylation. **(a),** Masses of methylated sugars are in parentheses and the ion peaks correspond to the sodium adducts of the methylated sugars. **(a-c)** relative intensity (% intensity) is shown.

A domain with a previously undescribed fold is required for the activity of GH136 enzymes on HMOs

216
217

218 The identification of a RhLnb136$_{II}$ homolog with high (>60 %) amino acid sequence identity in *E.*
219 *ramulus* (*Er*Lnb136; Supplementary Fig. 1 and 8a,b) is consistent with the proficient growth of this
220 strain on LNT. Surprisingly, no adjacent GH136$_I$ homolog was identified. Instead, the functionally
221 unassigned N-terminal region of *Er*Lnb136 displayed ≈40% amino acid sequence identity to
222 RhLnb136$_I$ (Supplementary Fig. 8b), which was suggestive of the fusion of *Er*Lnb136$_I$ and
223 *Er*Lnb136$_{II}$ into a single protein. Indeed, *Er*Lnb136 displayed ≈3.9 fold higher catalytic efficiency
224 on LNT than RhLnb136 (Supplementary Table 4). A single thermal transition was observed for
225 the unfolding of *Er*Lnb136, suggesting a cooperative unfolding and intimate interaction of the two
226 domains (Supplementary Fig. 8c). We crystallized this enzyme to discern the interactions between
227 the two subunits/domains compulsory for activity within GH136. The crystal structures of
228 selenomethionine (SeMet)-labelled and native *Er*Lnb136 were determined at 1.4 Å and 2.0 Å
229 resolution, respectively (Supplementary Table 6). The C-terminal catalytic GH136 domain
230 (*Er*Lnb136$_{II}$, from AA 242-663) assumes a β helix fold (Fig. 3) similar to the bifidobacterial homolog
231 BlLnbX (Supplementary Table 7). The LNB molecule bound in the active site is recognized by ten
232 potential hydrogen bonds and aromatic stacking of the Gal unit onto W548 (Fig. 3f and
233 Supplementary Fig. 9a). Interestingly, the GlcNAc sugar ring of LNB in *Er*Lnb136 adopts an ^4E
234 conformation (φ = 232° and ψ = 68°), enabling the O1-OH to adopt a pseudo-axial position to
235 form a direct hydrogen bond with the catalytic acid/base residue (D568) (Supplementary Fig. 9a).
236 Moreover, the D575 O$^{\delta2}$ of the nucleophile is positioned appropriately for a nucleophilic attack on
237 the anomeric carbon of the GlcNAc at a distance of 3.2 Å (Fig. 3f).

238 The N-terminal domain (*Er*Lnb136$_I$, from AA 7-224) consists of 8 α-helices (α1-α8) (Fig. 3a-c) and
239 assumes a previously unknown fold, stabilized by the central helix α1. The structurally most
240 related protein to *Er*Lnb136$_I$, a peptidyl-prolyl cis-trans isomerase with a chaperone activity from
241 *Helicobacter pylori* (5EZ1), shares weak structural similarity restricted to helices α6 and α7
242 (Supplementary Fig. 9b, Supplementary Table 7). The *Er*Lnb136$_I$ domain embraces the sides and
243 back of the β helix domain (Fig. 3a-c). These extensive inter-domain interactions (solvent
244 inaccessible interface ≈1618 Å^2), stabilize the protein structure with ΔG= -17 kcal mol^{-1}.
245 Remarkably, the α6-α7 loop of *Er*Lnb136$_I$ forms a part of the active site with the solvent accessible
246 sidechain of Y145 positioned near the active site (5.7 Å to the GlcNAc O1 atom of LNB) (Fig. 3d,
247 e). The Y145A mutant showed a 4.9-fold higher K_M (Supplementary Table 4, Supplementary Fig.
248 8d), suggesting that this residue contributes to substrate interactions, possibly at the +1 subsite.

249
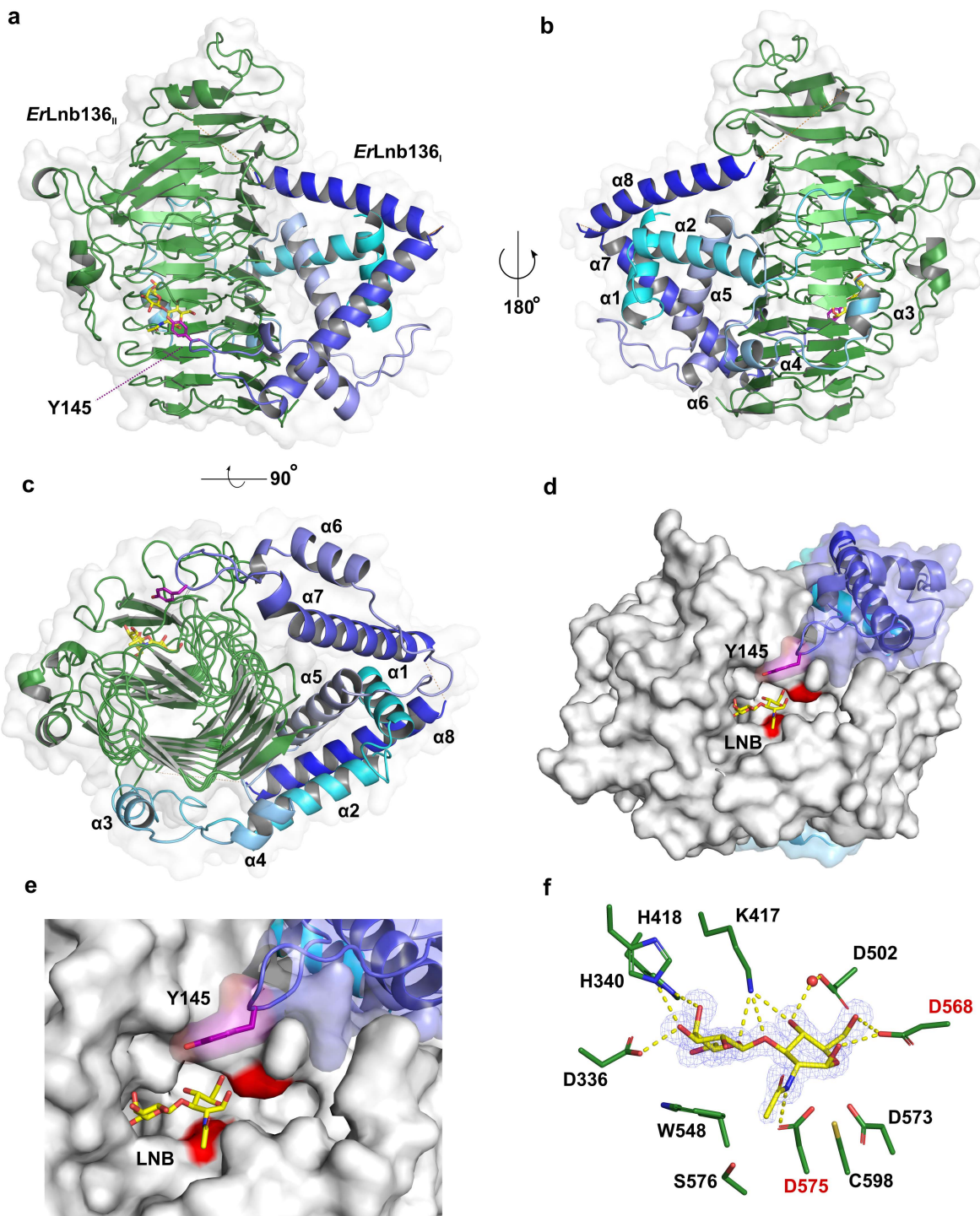
Fig. 3: Crystal structure of the GH136 lacto-*N*-biosidase from *E. ramulus* (*Er*Lnb136). (a-c) Overall structure and a semitransparent surface of *Er*Lnb136 consisting of an N-terminal domain designated as *Er*Lnb136ᵢ (cyan-blue) and a C-terminal β-helix domain (green) -*Er*Lnb136ᵢᵢ. The enzyme is shown in **(a)** a view orthogonal to the C-terminal β helix domain, **(b)** the view of (a) rotated 180˚ and **(c)** a view along the axis of C-terminal β helix domain, to highlight the interaction of the *Er*Lnb136ᵢ and *Er*Lnb136ᵢᵢ domains. **(d)**, A molecular surface top view of the active site and a close up view **(e)** to illustrate the contribution of the N-terminal *Er*Lnb136ᵢ domain to the active site architecture, especially the tyrosine (Y145, magenta) that contributes to substrate affinity. **(f)** The weighted mF_o-DF_c omit electron density map (contoured at 4.0 σ) of the LNB unit (yellow sticks) bound at the active site of *Er*Lnb136 is shown. The water (red sphere) mediated and direct hydrogen bonds that recognize the LNB are shown as yellow dashed lines. **(d-f)** The catalytic nucleophile (D575) and catalytic acid/base residue (D568) are highlighted with red labels. **(a-c)** Disordered regions (residues 180-199 and 225-241) are shown as orange dotted lines

Cross-feeding and increased butyrate production from *Roseburia* in mucin cocultures with *Akkermansia muciniphila*

HMOs and *O*-glycans from glycolipids and glyco-proteins including mucin share structural motifs. The high affinity of the SBPs from *Roseburia* for GNB suggested possible foraging of mucin (and/or oligomers from glycoconjugates) and thereby a metabolic interplay of *Roseburia* with mucolytic HGM members. To evaluate possible mechanisms of cross-feeding we compared *Roseburia* growth on mucin with and without the model mucin degrader *A. muciniphila*[30].

A co-culture of *R. hominis* and *R. inulinivorans* displayed no growth within 24 h on a mucin mixture and only poor growth after 48 h (Supplementary Fig. 7a,b), in contrast to *A. muciniphila* that grew well within 24 h. The co-culture of the two *Roseburia* species and *A. muciniphila* grew to a significantly higher OD_{600} than *A. muciniphila* alone ($p < 3.7 \times 10^{-6}$ at 24 h, $p < 1.3 \times 10^{-3}$ at 48 h)(Supplementary Fig. 7a). This growth is supported by a 4.5 fold higher butyrate level in the co-culture supernatants than *Roseburia* alone (24 h). After 48 h, a slight increase in butyrate concentration was also detected in cultures containing only *Roseburia* consistent with the growth data (Supplementary Fig. 7c).
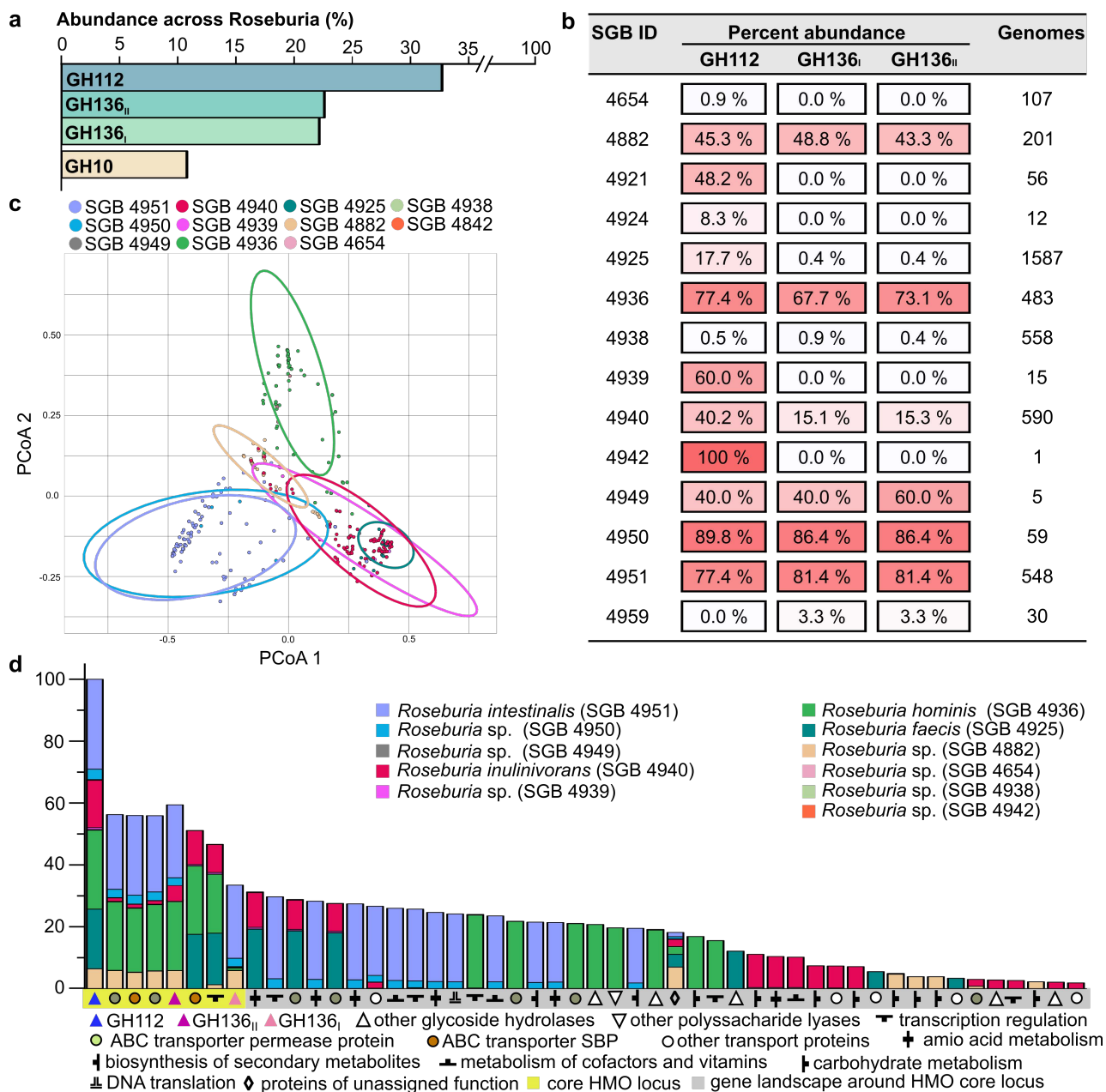
To unveil the basis for the *Roseburia* growth, the proteomes of *R. hominis* and *R. inulinivorans* were compared between co-cultures of *Roseburia* and *A. muciniphila* grown on mucin and glucose, respectively. For *R. hominis* and *R. inulinivorans*, 31 and 93 proteins, including several CAZymes, were significantly upregulated (log_2 fold change > 2) (Supplementary Fig. 7e-h) relative to the glucose co-cultures. The transport protein *Rh*LNBBP and *Rh*GLnbp112 from the *R. hominis* HMO locus (Fig, 1f) were the top 6th and 10th most upregulated proteins in the mucin proteome of *R. hominis*, respectively, highlighting the role of this locus in cross-feeding on host glycans (Supplementary Fig. 7g). In *R. inulinivorans,* the corresponding proteins *Ri*Le$^{a/b}$BP and *Ri*LNBBP were also significantly upregulated with log_2 fold changes of 2.77 and 4.74, respectively. However, the top upregulated protein in the *R. inulinivorans* proteome was a SBP of an ABC transporter co-localised with genes encoding a blood group A- and B- cleaving endo-β-(1→4)-galactosidase (*Ri*GH98), a putative α-galactosidase of GH36 and an α-L-fucosidase (GH29), which was the top fourth upregulated protein in the mucin proteome of *R. inulinivorans* (Supplementary Fig. 7h)*.* The upregulation of this locus suggested that *R. inulinivorans* possesses a functional machinery for directly accessing certain mucin oligomers. We expressed the predicted extracellular *Ri*GH98 and demonstrated robust release of blood group A and B oligomers from mucin and related *O*-glycans (Supplementary Fig. 7i-j, Supplementary Table 9). The co-upregulation of a locus encoding a fucose utilization pathway (Supplementary Fig. 3a) is in accordance with the release of fucosylated oligomers by *Ri*GH98. Another route of foraging, was suggested by the high upregulation of the sialic acid catabolism pathway (Supplementary Fig. 3b), which likely confers the potent growth of *R. inulinivorans* on this substrate (Fig. 1d). These findings establish that the HMOs utilization machinery and additional functional operons support co-growth with *A. muciniphila* on mucin.

The HMO utilization loci are preponderant in the *Roseburia* genome

The HMO loci, defined by the co-occurrence of GH136 and GH112 genes, are conserved in 5 *Roseburia* reference genomes (Supplementary Fig. 1). To broadly examine the structure and conservation of these loci, the presence of homologs of the aforementioned genes was mapped across 4599 previously reconstructed *Roseburia* genomes[31]. As a reference signature for a central catabolic pathway, the presence of GH10 xylanase genes, compulsory for xylan utilization as shown in *R. intestinalis*[32], was also analyzed. Strikingly, the GH112 and GH136 HMO utilization

307   genes are about 2-3 fold more abundant than the GH10 counterparts (Fig. 4a), indicative of the
308   broader distribution of the HMO loci compared to the xylanase locus, which is mainly conserved
309   in *R. intestinalis*. The GH136$_I$ and GH136$_{II}$ genes have a similar abundance, which is about 30 %
310   lower than that of GH112. This overall trend is reiterated from analyses of individual species-level
311   genome bins (SGBs), but differences in the co-occurrence patterns of GH136 and GH112 genes
312   in different *Roseburia* phylotypes were observed (Fig. 4b). Interestingly, GH136 genes were either
313   absent or far less abundant than GH112 counterparts in several SGB, whereas the converse was
314   only true for a single SGB (4959) that represents 0.65% of the total number of the analysed
315   *Roseburia* genomes. The higher abundance of GH112 genes, associated with the phosphorolysis
316   of HMO or mucin derived disaccharides, prompted us to analyze the organization of 1397 loci,
317   defined by the presence of GH112 sequences in the same metagenomic dataset with a more
318   stringent threshold (70% identity to GH112 sequences present in 5 *Roseburia* reference genomes
319   (Supplementary Fig. 1)). The composition of the gene landscapes appeared to be SGBs specific
320   (Fig. 4c). An ABC transporter, GH136$_I$/GH136$_{II}$, and transcriptional regulators were, however, the
321   most frequently co-occurring genes with GH112 gene, which offers a robust signature of the core
322   HMO utilization loci (Fig. 4d) and validates the broad distribution of the pathways described in the
323   present study. Additional CAZymes and carbohydrate metabolic genes were also frequently co-
324   occuring in the vicinity of GH112 genes, suggesting that additional glycan utilization capabilities
325   are clustered around the HMO loci. The *Ri*GH136-like sequences (Fig. 4d) are likely to be
326   underestimated due to the divergence of this clade of GH136 that resembles a previously
327   unknown specificity.

328

a Abundance across Roseburia (%)

b

SGB ID	Percent abundance			Genomes
	GH112	GH136$_I$	GH136$_{II}$	
4654	0.9 %	0.0 %	0.0 %	107
4882	45.3 %	48.8 %	43.3 %	201
4921	48.2 %	0.0 %	0.0 %	56
4924	8.3 %	0.0 %	0.0 %	12
4925	17.7 %	0.4 %	0.4 %	1587
4936	77.4 %	67.7 %	73.1 %	483
4938	0.5 %	0.9 %	0.4 %	558
4939	60.0 %	0.0 %	0.0 %	15
4940	40.2 %	15.1 %	15.3 %	590
4942	100 %	0.0 %	0.0 %	1
4949	40.0 %	40.0 %	60.0 %	5
4950	89.8 %	86.4 %	86.4 %	59
4951	77.4 %	81.4 %	81.4 %	548
4959	0.0 %	3.3 %	3.3 %	30

c

d

329
330

Fig. 4: The conservation and structure of HMO utilization loci in *Roseburia*. (a) Global abundance of GH112, GH136$_I$, GH136$_{II}$ and GH10 xylanase genes in 4599 *Roseburia* genomes. **(b)** Heatmap showing the segregation of GH112 containing genomes from (a) into different species-level genome bins (SGBs) and the corresponding relative abundance patterns of HMO utilization proteins within each SGB. **(c)** Principal component analysis of 1397 *Roseburia* gene landscapes defined stringently based on >70% identity to the GH112 with any of the 5 references *Roseburia* genomes (Supplementary Fig. 1) and including 10 proteins up and downstream of the GH112. **(d)** The composition of gene landscapes defined in (c) viewed as the occurrence frequency relative to GH112 genes. The gene landscape analyses **(c-d)** were performed using an automatic annotation pipeline, which differentiated between two different ABC-transporter solute binding proteins (SBPs) in the core HMO locus. The *R. inulinivorans* GH136 sequences are likely to be under estimated due to their divergence from canonical lacto-*N*-biosidase counterparts e.g. from *Roseburia hominis* characterized in this study. The HMO utilization loci are defined by genes encoding at least one GH112, a GH136, an ABC-transporter, and a transcriptional regulator.

Discussion

Perturbation of the early life HGM assembly is associated with life-long effects on the immune- and metabolic homeostasis of the host[9–12]. Breastfeeding is a key affector of the dynamics of the microbiota during infancy. Weaning marks a dramatic transition towards an adult-like structure of the HGM, which matures at the age of 2-3 and exhibits high resilience throughout adulthood[7,8,22].

The critical window that precedes the maturation of the microbiota offers a unique opportunity for therapeutic interventions to address aberrant HGM states and thereby to prevent dysbiosis-related chronic disorders. To date, insight into the compositional signatures that characterize the assembly of the microbiota during infancy[6–8] is available, but the underpinning mechanisms, especially during weaning, remain largely unknown. Here, we describe previously unknown pathways that confer the growth of butyrate producing Clostridiales on distinct HMO motifs and related oligomers from host glyco-conjugates. These pathways correlate to the early colonization by Clostridiales associated with the healthy HGM and with the protection from metabolic and inflammatory disorders as well as colorectal cancer[24–26,33].

The protein apparatus that confers the metabolism of HMO motifs and related glyco-conjugate oligomers in butyrate producing Firmicutes of the HGM

We uniquely demonstrate that key butyrate producing *Roseburia* and *Eubacterium* spp. grow on complex HMOs purified from mother's milk and on defined HMO motifs (Fig. 1a-c). Proteomic analyses revealed two highly upregulated genetic loci that encode distant homologs to a lacto-*N*-biosidase from *B. longum* [28,29], GNB/LNB phosphorylases and ABC transporters in *R. hominis* and *R. inulinivorans*, (Fig. 1f-g and Supplementary Fig. 1). Our analyses (Fig.1 and 2, Supplementary Fig. 6, Table 1, Supplementary Tables 2, 3, 4 and 5) established that the locus of *R. hominis* supports growth on the HMO motifs LNT and LNB, whereas the *R. inulinivorans* locus confers growth on more complex HMOs, e.g. single and double fucosylated versions of LNT (Fig. 2a and Supplementary Fig. 6a-d). This specialization on different, but partially overlapping, HMOs and related Lewis a and Lewis b antigen oligomers from glyco-lipids or glyco-proteins creates differential competitive catabolic niches. This specialization is evident from the divergence of the GH136 specificities. Thus, *Rh*Lnb136 and *Er*Lnb136 are lacto-*N*-biosidases, whereas *Ri*Le$^{a/b}$136 displays an unprecedented specificity that requires a Fuc-α-(1→4)-GlcNAc at the proximal glycone subsite (subsite -1) and accommodates additional fucosylation at the -2, and +2 subsites (Fig. 2, Supplementary Fig. 6a and Supplementary Table 4). The preference to fucosylation is consistent with an open active site effectuated by shortening of loops, (*Er*Lnb136: Loop 1 AA 330-341, Loop 2 AA 520-543, Supplementary Fig. 9c), which allows the accommodation of bulky fucosylated substrates. Remarkably, the GH136$_I$ subunits (or domains in *Er*GH136-like enzymes) are co-evolved with the GH136$_{II}$ counterparts that possess the catalytic residues (Supplementary Fig. 9d).

Our stability (Supplementary Fig. 8c), structural (Fig. 3 and Supplementary Fig. 9), biochemical (Supplementary Fig. 8, Supplementary Table 4) and phylogenetic analyses (Supplementary Fig. 9d) affirm the crucial role of the GH136$_I$ domain in the functionality of GH136 enzymes and provide the first insight into the association of the two GH136 domains. The sequence conservation of GH136$_I$ and GH136$_{II}$ was mapped on the structure of *Er*Lnb136. Strikingly, highly conserved patches were identified across both domains (Supplementary Fig. 9e). Particularly, parts of the α4-α5 loop and of the α5 helix in *Er*Lnb136$_I$ that pack extensively onto *Er*Lnb136$_{II}$ display globally conserved residues, together with the complementary co-conserved regions of *Er*Lnb136$_{II}$

387 (Supplementary Fig. 9e). Moreover, the surface of $ErLnb136_I$ is positively charged and apolar at
388 the interface with $ErLnb136_{II}$, which is notably different from the negative potential on the surface
389 of the rest of the enzyme (Supplementary Fig. 9f) and complementary to the interface surface of
390 $ErLnb136_{II}$. These results highlight the co-evolution of GH136 subunits or domains.

391 ABC transporters are a determinant of uptake selectivity and competitiveness in both
392 bifidobacteria[17,34,35] and *Roseburia intestinalis*[32]. The two SBPs of the ABC importers located in
393 the HMO loci of *R. hominis* and *R. inulinivorans* were within the top 5 upregulated proteins in each
394 proteome in response to HMO utilization (Fig. 1), underscoring the critical role of oligosaccharide
395 transport in the competitive gut niche. The preferences of the SBPs and GHs encoded by these
396 loci appear aligned to confer efficient uptake and subsequent catabolism of preferred substrates
397 (Fig. 2 and Supplementary Fig. 6, Table 1). The LNB/GNB phosphorylases of GH112 are also
398 conserved in the HMO loci (Supplementary Fig. 1). *R. inulinivorans* possesses additional
399 CAZymes, notably different fucosidases for degradation of internalized fucosylated-oligomers
400 (Supplementary Fig. 1 and 6b-d). Based on the proteomic analyses and the biochemical data,
401 we propose a model for the two distinct routes for uptake and depolymerisation of HMOs in
402 *Roseburia* and *Eubacterium* (Fig. 5 and Supplementary Fig. 2).

403 Conserved HMO utilization loci correlate to early colonization and resilience of
404 Clostridiales from *Roseburia* and *Eubacterium*

405 Butyrate producing bacteria of the *Roseburia-Eubacterium* group (Clostridiales order) are early
406 colonizers of the infant gut[6,8,36] and are prevalent members of the adult HGM[22,23].

407 The origin of this taxonomic group is enigmatic, but their presence in the human milk microbiome
408 has been reported[37,38]. Orthogonal transfer from mothers based on the identification of the same
409 *Roseburia* strains in mothers faeces, milk and the infant guts[39] has also been proposed. *R.*
410 *intestinalis* type strains have been isolated from infant faeces[40], hinting the presence of this taxon
411 before full transition to solid food. The pathways we have elucidated explain, at least partially, the
412 association between HMO utilization and the early colonization of butyrate producing *Roseburia*
413 and *Eubacterium*.

414 We have previously shown that the abundance of distinct bifidobacteria in guts of breast-fed
415 infants is strongly correlated to efficient ABC transporters that capture the 2'- and 3'-fucosyl-
416 lactose HMOs with high affinity ($K_D \approx 5$ μM)[17]. The strains possessing these genes, e.g. from
417 *Bifidobacterium longum* subspecies *infantis*, are not detected after weaning, as opposed to
418 counterparts adept at utilizing plant-derived glycans. By contrast, the same *Clostridium* group
419 XIVa strains that possess plant glycan utilization pathways[32,41,42] retain HMO catabolic pathways.
420 The simultaneous growth of *R. hominis* on LNT and the plant derived xylotetraose
421 (Supplementary Fig. 5) demonstrates this catabolic plasticity, which likely confers competitive
422 advantages during weaning.

423 The loci that target HMOs also mediate cross-feeding on mucin or other glyco-conjugate
424 oligomers, e.g. GNB from mucin and blood antigen structures, both captured efficiently by
425 *Roseburia* transport proteins (Table 1). This is consistent with the significant butyrate production
426 measured in co-cultures of *Roseburia* and *A. muciniphila*[30] (Supplementary Fig. 7c) and the
427 upregulation of GH136-containing loci in the mucin co-culture and HMO monocultures (Fig 1 and
428 Supplementary Fig. 7g). *R. inulinivorans* possesses an extensive mucolytic machinery revealed
429 by the upregulation a fucose, a sialic acid (Supplementary Fig. 3) and a blood group A and B-
430 locus (Supplementary Fig. 7h-j, Supplementary Table 9) that allows the release of β-(1→4)-linked

431   blood group oligomers found in mucin and glyco-lipids on the surfaces of enterocytes[43,44]. This
432   ability to access carbohydrates from mucin and host glyco-conjugates supports growth during
433   periods of nutritional perturbations, which may increase the resilience of this taxonomic group.

434   Our metagenomics analyses establish that HMO utilization appears to be a core trait within
435   *Roseburia*, based on the ubiquitousness of loci harboring GH112 and GH136 genes (Fig. 4). The
436   presence of SGBs that exclusively possess GH112 genes (e.g. SGBs 4921 and 4939, Fig. 4b)
437   suggests that distinct strains are secondary degraders that cross-feed on released simple
438   substrates, e.g. LNB and GNB. By contrast, the balanced occurrence of GH112 and GH136 genes
439   (Fig. 4b) offers a signature for primary degraders that are able to access more complex glycans
440   from HMOs or host glyco-conjugates.

441   In conclusion, the present study sets the stage for a mechanistic understanding of the assembly
442   of core physiologically important groups in the early life microbiota and discloses previously
443   unknown roles of HMOs in selection of Clostridiales. Additional studies are required to further
444   address the paramount, but poorly understood maturation of the early life microbiota.

445
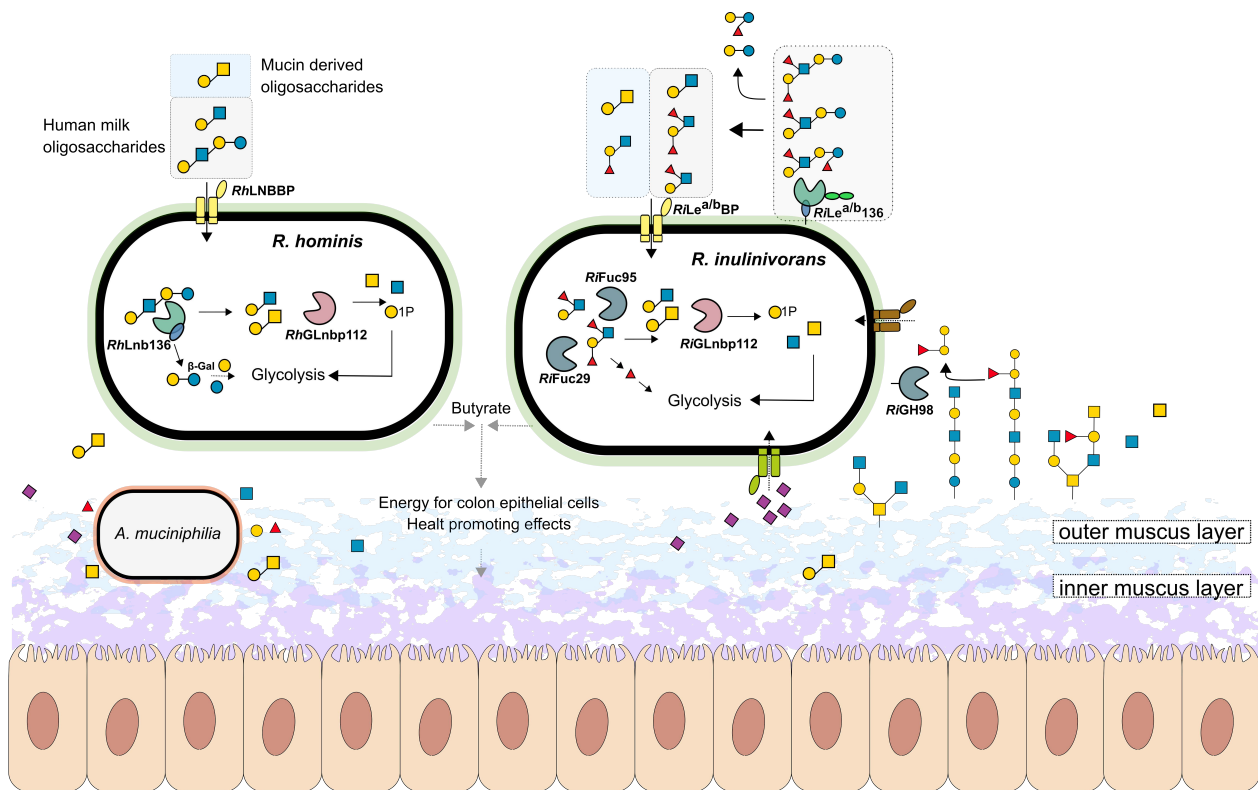
446

447   **Fig. 5: Model for HMOs and related host glycan utilization by *Roseburia* and other Lachnospiraceae.** In *R.*
448   *hominis* the HMOs blocks LNT and LNB and the mucin derived GNB are captured by *Rh*LNBBP for uptake into the
449   cytoplasm and LNT is subsequently hydrolyzed to LNB. Both LNB and GNB are phosphorolyzed by *Rh*GLnbp112 into
450   α-D-galactose-1-phosphate and the corresponding *N*-acetylhexosamines GlcNAc and GalNAc, respectively. Lactose is
451   likely hydrolysed by a canonical β-galactosidase. *R. inulinivorans* specializes on different HMO blocks and structurally
452   overlapping oligomers from glyco-lipids or glyco-proteins. Initial hydrolysis of HMOs or *O*-glycans from glyco-
453   lipids/proteins occurs at the outer cell surface of *R. inulinivorans* by the activity of *Ri*Le[a/b]136, which has two C-terminal
454   putative galactose-binding domains. The capture and import of degradation products is mediated by *Ri*Le[a/b]BP and the
455   associated ABC transporter. In the cytoplasm, fucosyl decorations are remove by the concerted activity of *Ri*Fuc95 and

456   *Ri*Fuc29 before *Rh*GLnbp112 phosphorolyzes the resulting LNB or imported GNB into monosaccharides, as described
457   in *R. hominis*. The galactose and galactose-1-phosphate products are converted via the Leloir pathway to glucose-6-
458   phosphate and *N*-acetylhexosamine sugars are converted to GlcNAc-6-phosphate before entering glycolysis. The
459   pyruvate generated from glycolysis is partly converted to butyrate[45]. To make the model more appropriate *Roseburia*
460   is presented in its ecological niche, the outer mucus layer[46], together with *Akkermansia muciniphilia* as model mucin
461   degrader. The ability of *R. inulinivorans* to cross-feed on sialic acid and to directly access host glycans is illustrated by
462   the presence of the sialic acid uptake and catabolism machinery and by *Ri*GH98, cleaving β-(1→4)-linked blood group
463   A and B oligosaccharides from mucin and glyco- lipids/proteins. Black solid arrows show enzymatic steps established
464   or confirmed in this study. Black dotted arrows indicate steps based on literature. Grey dotted arrows indicate butyrate
465   production by *R. hominis* and *R. inulinivorans* from mucin in co-culture with *A. muciniphilia*. The glycan structure key is
466   the same as in Fig. 1.

467

468

469

470

471

472

Material and Methods:

474

Resources: Chemicals and Carbohydrates:

476

Human milk and blood antigen oligosaccharides used in this study are described in Table S1. *N*-acetylneuraminic acid (Neu5Ac), α-D-galactose-1-phosphate (Gal1P) and α-L-fucose (Fuc) were form Carbosynth and xylotetraose was from Megazyme. Galactose (Gal), Glucose (Glc), *N*-acetylglucosamine (GlcNAc), *N*-acetylgalactosamine (GalNAc) and porcine gastric mucin type III, (PGM) were from Sigma Aldrich. Bovine submaxillary mucin (BSM) was from VWR. 2-aminoanthranilic acid (2-AA) was from Nacalai Tesque and purified human milk samples were prepared from mother milk from Hvidøvre hospital (Hvidøvre, Denmark). All chemicals were of analytical grade unless otherwise stated.

Method Details:

Enzymatic production of LNB and GNB

LNB and GNB for growth were produced enzymatically with the GH112 galacto-*N*-biose/lacto-*N*-biose phosphorylase (EC 2.4.1.211) from *R. hominis* (*Rh*GLnbp112). In detail, 100 mM Gal1P and 300 mM corresponding *N*-acetylhexosamine (GlcNAc or GalNac) in 50 mM MES, 150 mM NaCl, pH 6.5 were incubated with 10 µM *Rh*GLnbp112 for 36 h at 30°C. After incubation, 2.5 volumes of ice-cold ethanol (99 %) were added, samples were incubated at – 20°C for 2 h and centrifuged (10.000x g, 30 min at 4°C) to remove the enzyme. Supernatants were up concentrated by rotary evaporation and disaccharides were desalted in ultrapure water (milliQ) using a HiPrep Desalt column (GE Healthcare, Denmark) on an Äkta avant chromatograph (GE Healthcare). Elution was monitored by measuring $A_{235\ nm}$ and pooled fractions were freeze dried. Further purification was accomplished by high-performance liquid chromatography (HPLC) (UltiMate 3000, Dionex) using a TSKgel® Amide 80 column (4.6 x 250 mm) and a TSKgel® Amide 80 guard column (4.6 x 10 mm) (VWR) by loading LNB or GNB dissolved in the mobile phase (75% (v/v) acetonitrile, ACN) and an isocratic elution at 1 mL min^{-1}. Purity of collected fractions (2 mL) was analyzed by thin layer chromatography (TLC) using 5 mM standards of GalNAc, GlcNAc, Gal1P and LNB/GNB. Fractions containing pure LNB/GNB were pooled, ACN was removed by speed vacuum evaporation and samples were lyophilized until further use.

Purification of human milk oligosaccharides

Human milk oligosaccharides (HMOs) were purified from pooled human milk samples as previously described[47,48]. Milk fat was separated by centrifugation (10.000x g, 30 min at 4°C) and proteins were removed by ethanol precipitation (as above). The supernatant was up concentrated by rotary evaporation, buffered with 2 volumes 100 mM MES, 300 mM NaCl, pH 6.5 and lactose was digested with ß-galactosidase from *Kluyvermomyces lactis* (Sigma Aldrich) (20 U mL^{-1}, 3 h at 37°C). The enzyme was precipitated with ethanol (as before) and the supernatant was concentrated by rotary evaporation. Residual lactose and monosaccharides were removed by solid-phase extraction (SPE) using 12 mL graphitized Supelclean™ ENVI-Carb™ columns (Supelco) with a bed weight of 1 g. For SPE, columns were activated with 80% (v/v) ACN containing 0.05% (w/v) formic acid (FA) and equilibrated with buffer A (with 4% (v/v) ACN, 0.05%

514   (w/v) FA), which was also used to dilute the samples prior to loading. After sample loading, the
515   columns were washed (6 column volumes of buffer A) to remove lactose and monosaccharides
516   before oligosaccharides were eluted with 40% (v/v) ACN, 0.05% (w/v) FA. Eluted
517   oligosaccharides were concentrated in a speed vacuum concentrator, freeze-dried and dissolved
518   in milliQ prior to usage.
519   Purity of HMOs was verified by high performance anion exchange chromatography with pulsed
520   amperometric detection (HPAEC-PAD) on an ICS-5000 (Dionex) system with a 3×250 mm
521   CarboPac PA200 column (Theromofisher), a 3×50 mm CarboPac guard column
522   (Theromofisher) and 10 µL injections. HMOs were eluted with a stepwise linear gradient of
523   sodium acetate: 0-7.5 min of 0-50 mM, 7.5-25 min of 50-150 mM and 25-35 min of 150-400
524   mM, at a flow rate of 0.35 mL min^{-1} and a mobile phase of constant 0.1 mM NaOH. Standards
525   (0.01-0.5 mM) of lactose, galactose and glucose in milliQ were used to quantify these residual
526   sugars as described above. The analysis was performed in triplicates and the residual content
527   of these sugars was <2 % (w/w) of the purified HMO mixture.

Isolation and purification of porcine mucins

529   The commercial porcine gastric mucin (PGM) was further purified[49]. In short, 20 g PGM was
530   stirred for 20 h at 25°C in 20 mM phosphate buffer, 100 mM NaCl, pH 7.8 (adjusted to pH 7.2
531   after the first 2 hours using 2 M NaOH). Insoluble residues were removed by centrifugation
532   (10,000x g, 30 min at 4°C) and soluble mucin was precipitated by the addition of 3 volumes of ice
533   cold ethanol (99%) and incubation for 18 h at 4°C. Precipitated mucin was dialyzed 5 times against
534   200 volumes milliQ for 16 h at 4°C, using a 50 kDa molecular weight cut off membrane (Spectra,
535   VWR) and afterwards freeze dried.

536   Porcine colonic mucin was isolated from five fresh pig colons from the slaughterhouse of Danish
537   Crown (Horsens, Denmark). Pig colons were processed at site and immediately placed on dry ice
538   to ensure quick cooling during transport. Colons were opened longitudinally and content was
539   removed mechanically and by washing with ice cold 0.9% (w/v) NaCl until no digesta was visible.
540   Cleaned luminal surface was quickly dried with absorptive paper and the mucosa was scraped
541   off with a blunt metal spatula and subsequently transferred into a pre-cooled glass beaker
542   whereby visible fat was removed and discarded. Mucin was then purified as previously
543   described[50]. Isolated mucin was immersed in 10 volumes extraction buffer (10 mM sodium
544   phosphate buffer, 6 M guanidine hydrochloride (GuHCl), 5 mM Ethylenediaminetetraacetic acid
545   (EDTA), 5 mM N-ethylmaleimide, pH 6.5) and gently stirred overnight at 4°C. Soluble impurities
546   and floating fat were separated by centrifugation (10,000x g, 30 min at 4°C), pelleted mucin was
547   dissolved in 10 volumes extraction buffer and incubated for 3 h at room temperature again.
548   Soluble impurities were removed by centrifugation as described before. Short incubation (3 h)
549   extraction steps were repeated 7 times until the supernatant was clear for at least two repeated
550   extractions. Afterwards insoluble mucin was solubilized by reduction in 0.1 M Tris, 6 M GuHCl,
551   5mM EDTA, 25 mM dithiotreitol (DTT) pH 8, for 5 h at 37°C and subsequent alkylation through
552   the addition of 65 mM iodoacetamide and incubation in the dark for 18 h at 4°C. Soluble mucin
553   was dialyzed 6 times against 200 volumes milliQ using a 50 kDa MWCO dialysis bag for 6 h at
554   4°C and freeze dried.

Cloning, expression and purification of proteins

556   Open reading frames encoding proteins from *R. hominis* DSM 16839, *R. inulinivorans* DSM 16841
557   and *E. ramulus* DSM 15684 were cloned without signal peptide or transmembrane domain from

558 genomic DNA using In-Fusion cloning (Takara) and the primers in Table S8 into the EcoRI and
559 NcoI restriction sites of the corresponding plasmids, to encode proteins with either a cleavable N-
560 or C- terminal His$_6$ tag. The pETM 11 plasmid was used (from G. Stier, EMBL, Center for
561 Biochemistry, Heidelberg, Germany)[51], except for *RHOM_04110* (*Rh*Lnb136$_I$) and
562 *ROSEINA2194_01899* (*Ri*Le$^{a/b}$136$_I$) which were cloned into pET15b (Novagen). Recombinant
563 proteins were expressed in *E. coli* BL21 Δ*lacZ* (DE3)/pRARE2 (a kind gift from Prof. Takane
564 Katayama, Kyoto University, Kyoto, Japan) and purified following standard protocols using His-
565 affinity and size-exclusion chromatography. Mutants of *E. ramulus HMPREF0373_02965*
566 (*Er*Lnb136) were constructed using QuickChange II Site-Directed Mutagenesis (Agilent) with
567 pETM11_ *HMPREF0373_02965* as template. Primers used for site-directed mutagenesis are
568 listed in Table S8 and mutants were produced as described above. L-Selenomethionine labelled
569 protein expression of *Er*Lnb136 was performed by introducing the corresponding plasmid into *E.*
570 *coli* B834 (DE3) and culturing the transformed cells in a synthetic M9 based medium of the
571 SelenoMet labelling Kit (Molecular Dimensions) supplemented either with L-methionine or L-
572 Selenomethionine (both to 50 µg mL^{-1}). The L-SeMet labelled protein was purified as described
573 above.
574
575
Growth experiments and single strain proteomics analysis

577 *R. hominis* DSM 16839, *R. inulinivorans* DSM 16841, *E. ramulus* DSM 15684 and *E. ramulus*
578 DSM 16296 were grown anaerobically at 37°C using a Whitley DG250 Anaerobic Workstations
579 (Don Whitley Scientific). *R. hominis* and *R. inulinivorans* were propagated in YCFA medium[40]
580 while for *E. ramulus* strains CFA medium (modified YCFA medium lacking yeast extract to
581 minimize *E. ramulus* growth on yeast extract) was used. Growth media were supplemented with
582 0.5% (w/v) carbohydrates sterilized by filtration (soluble carbohydrates, 0.45 µm filters) or
583 autoclaving (mucins, 15 min at 121°C) and cultures were performed in at least biological triplicates
584 unless otherwise indicated. Bacterial growth was monitored by measuring $OD_{600\ nm}$ and pH (for
585 co-culture experiments). For growth experiments performed in microtiterplates, a Tecan Infinite
586 F50 microplate reader (Tecan Group Ltd) located in the anaerobic workstation was used and
587 growth was followed by measuring $OD_{595\ nm}$.
588 For differential proteome analyses, *R. hominis* and *R. inulinivorans* were grown in 200 µL YCFA
589 (1.5 mL Eppendorf tubes) to mid-late exponential phase (OD_{600} ~0.5-0.8) in four biological
590 replicates. For *R. hominis* YCFA was supplemented with 0.5% (w/v) LNT or glucose and for *R.*
591 *inulinivorans* 1% (w/v) HMOs or glucose was used as carbon source. Cells were harvested by
592 centrifugation (5.000x *g*, 5 min at 4°C), washed twice with ice cold 0.9% (w/v) NaCl, resuspended
593 in 20 µL lysis buffer (50 mM HEPES, 6 M GuHCl, 10 mM Tris(2-carboxyethyl)phosphine
594 hydrochloride (TCEP), 40mM 2-chloroacetamide (CAA) pH 8.5) and stored at −80°C for
595 proteomics analysis.
596
Co-culture cross-feeding experiment proteomics analyses

598 *R. hominis*, *R. inulinivorans* and *A. muciniphila* DSM 22959 were grown in 10 mL YCFA to mid-
599 late exponential phase (OD_{600} ~0.6-0.7). From these pre-cultures, equal amounts of cells (OD_{600})
600 were used to inoculate 30 mL fresh YCFA medium with 1% (w/v) of a mucin mixture (0.6% (w/w)
601 PGM, 0.2% (w/w) PCM, 0.2% (w/w) BSM) or 1% (w/v) glucose to a start OD_{600} ~0.01. All cultures
602 were performed in four biological replicates and growth was followed (OD_{600} and pH) at 0, 6, 8,
603 12, 16, 24, and 48 h. Samples (2 mL) were collected for proteomics analyses after 16 h and for

604 SCFA quantification after 24 and 48 h. Samples were immediately cooled on ice and cells were
605 harvested by centrifugation (5000x *g*, 10 min at 4°C). For proteomics, cell pellets were washed
606 twice with ice cold 0.9% (w/v) NaCl, resuspended in 60 µL lysis buffer and stored at −80°C until
607 proteomics analysis. Collected culture supernatants for SCFA quantification were sterile filtrated
608 (0.45 µm filters) and stored at −80°C for further analysis.

609 **Protein extraction and sample preparation for mass spectrometry**

610 Samples were processed as described elsewhere[52,53]. Cells were lysed by boiling (5 min 95°C)
611 followed by bead beating (3mm beads, 30 Hz for 1 min) (TissueLyser II, Qiagen) and sonication
612 bath (3x 10 sec at 4°C) (Bioruptor, Diagenode). Lysates were centrifuged (14.000x *g*, 10 min at
613 4°C) and soluble protein concentrations were determined by a Bradford assay (Thermo Fisher
614 Scientific). For digestion, 20 µg protein were diluted 1:3 with 50 mM HEPES, 10% (v/v) ACN, pH
615 8.5 and incubated with LysC (MS grade, Wako) in a ratio of 1:50 (LysC:protein) for 4 h at 37°C.
616 Subsequently, samples were diluted to 1:10 with 50 mM HEPES, 10% (v/v) ACN, pH 8.5 and
617 further digested with trypsin (MS grade, Promega) in a ratio of 1:100 for 18 h at 37°C. Next,
618 samples were diluted 1:1 with 2% (w/v) trifluoroacetic acid (TFA) to quench enzymatic activity and
619 peptides were processed for mass spectrometry using in house packed stage tips[54] as described
620 below.
621 Peptides from single strain cultures were desalted using 3 discs of C18 resin packed into a 200
622 µL tip and activated by successive loading of 40 µL of MeOH and 40 µL of 80% (v/v) ACN, 0.1%
623 (w/v) FA by centrifugation at 1800x *g* and equilibrated twice with 40 µL of 3% (v/v) ACN, 1% (w/v)
624 FA before samples were loaded in steps of 50 µL. After loading, tips were washed three times
625 with 100 µL 0.1% (w/v) TFA and peptides were eluted in two steps with 40 µL each of 40% (v/v)
626 ACN, 0.1% (w/v) FA into a 0.5 mL Eppendorf LoBind tube. Peptides derived from co cultures were
627 desalted and fractionated using strong cation exchange (SCX) chromatography filter pugs (3M
628 Empore). Per sample, 6 SCX discs were packed into a 200 µL tip and tips were activated and
629 equilibrated by loading 80 µL (v/v) of ACN and then 80 µL of 0.2% (w/v) TFA. Samples were
630 applied in 50 µL steps and tips were washed twice with 600 µL 0.2% (w/v) TFA. Subsequently
631 peptides were stepwise eluted in 3 fractions with 60 µL of 125 mM NH_4OAc, 20% (v/v) ACN, 0.5%
632 (w/v) FA, then with 60 µL of 225 mM NH_4OAc, 20% (v/v) ACN, 0.5% (w/v) FA and lastly with 5%
633 (v/v) NH_4OH, 80 % (v/v) ACN into 0.5 mL Eppendorf LoBind tubes. Eluted peptides were dried in
634 an Eppendorf Speedvac (3 h at 60°C) and reconstituted in 2% (v/v) ACN, 1% (w/v) TFA prior to
635 mass spectrometry (MS) analysis.

636 **LC-MS/MS**

637 Peptides from biological triplicates of each culture condition were loaded on the mass
638 spectrometer by reverse phase chromatography through an inline 50 cm C18 column (Thermo
639 EasySpray ES803) connected to a 2 cm long C18 trap column (Thermo Fisher 164705) using a
640 Thermo EasyLc 1000 HPLC system. Peptides were eluted with a gradient of 4.8–48 % (v/v) ACN,
641 0.1% (w/v) FA at 250 nL min⁻¹ over 260 min (samples from single strain cultures) or 140 min (SCX
642 fractionated samples from co cultures) and analysed on a Q-Exactive instrument (Thermo Fisher
643 Scientific) run in a data-dependent manner using a "Top 10" method. Full MS spectra were
644 collected at 70,000 resolution, with an AGC target set to $3x10^6$ ions or maximum injection time of
645 20 ms. Peptides were fragmented via higher-energy collision dissociation (normalized collision
646 energy=25). The intensity threshold was set to $1.7x10^6$, dynamic exclusion to 60 s and ions with
647 a charge state <2 or unknown species were excluded. MS/MS spectra were acquired at a

648   resolution of 17,500, with an AGC target value of 1×10^6 ions or a maximum injection time of 60
649   ms. The scan range was limited from 300–1750 m/z.

650

Protein identification and Label-free quantification and relative abundance in co-culture communities

653   Proteome Discoverer versions 2.2 & 2.3 were used to process and analyze the raw MS data files
654   and label free quantification was enabled in the processing and consensus steps. The spectra
655   from single strains proteomics were matched against the proteome database of *R. hominis* DSM
656   16839 (ID: UP000008178) or *R. inulinivorans* DSM 16841 (ID: UP000003561) respectively, as
657   obtained from Uniprot. The spectra from co-culture experiments were searched against a
658   constructed database consisting of the reference proteomes of the two *Roseburia strains* (as
659   above) and *A. muciniphila* DSM 22959 (ID: UP000001031). For spectral searches, oxidation (M),
660   deamidation (N, Q) and N-terminal acetylation were specified as dynamic modifications and
661   cysteine carbamidomethylation was set as a static modification. Obtained results were filtered to
662   a 1% FDR and protein quantitation was done by using the built-in Minora Feature Detector. For
663   analysis of the label-free quantification data, proteins were considered present if at least 2 unique
664   peptides (as defined in Proteome Discoverer) were identified and proteins had to be identified in
665   at least 2 out of the 3 samples analyzed per culture condition with high confidence.

666   Relative bacterial abundance in co-cultures was estimated based on strain unique peptides
667   identified with Unipept version 4.0[55]. To exclude peptides shared between closely related strains
668   from the analyses, all peptide sequences quantified via Proteome Discoverer were imported into
669   the Unipept web server and analyzed with the settings "Equate I and L" and "Advanced missed
670   cleavage handling" activated. The normalized sum of intensities of the resulting taxonomically
671   distinctive peptides was then used for assessing relative abundances of each strain.

Butyrate quantification

673   Butyrate in culture supernatants was quantified by HPLC coupled to a refracting index detector
674   (RID) and diode array detector (DAD) on an Agilent HP 1100 system (Agilent). Standards of
675   butyric acid (0.09-50 mM) were prepared in 5 mM H_2SO_4 for peak identification and quantification.
676   Samples from 4 biological replicates were analysed by injecting 20 µL of standard or filtrated (0.45
677   µM filter) culture supernatant on a 7.8 x 300 mm Aminex HPX-87H column (Biorad) combined
678   with a 4.6 x 30 mm Cation H guard column (Biorad). Elution of was performed with a constant
679   flow rate of 0.6 mL min^{-1} and a mobile phase of 5 mM H_2SO_4. Standards were analysed as above
680   in technical triplicates.

Oligosaccharide uptake preference of *R. hominis*

682   *R. hominis* was grown anaerobically in 250 µL YCFA medium with 0.5% (w/v) of an equal mixture
683   of xylotetraose and LNT in biological triplicates. Samples (20 µL) were taken after 0, 3.5, 5.5,
684   6.5, 8, 9.5 and 24h, diluted 10-fold in ice cold 100 mM NaOH and centrifuged (10 min at 5000x
685   *g* at 4°C) before supernatants were stored at –20°C until the HPAEC-PAD analysis. Standards
686   of 0.5 mM xylotetraose and LNT were prepared in 100 mM NaOH and used to identify
687   corresponding peaks in the chromatograms. Samples or standard were injected (2 µL
688   injections) on a 4 × 250 mm CarboPac PA10 column with a 4 × 50 mm CarboPac guard column

689   and eluted isocratically (0.750 mL min^{-1}, 100 mM NaOH, 10mM NaOAc). The analysis was
690   performed from a biological triplicate and standards were analyzed in technical duplicates.

691

Enzyme activity assays

693   Enzymatic activity assays were carried out in 50 mM MES, 150 mM NaCl, 0.005% (v/v) Triton X-
694   100, pH 6.5 standard assay buffer and in triplicates unless otherwise stated.

695   Hydrolysis kinetics and specific activities of the GH136 lacto-*N*-biosidases were measured using
696   a coupled enzymatic assay to monitor lactose release. The lactose was hydrolyzed with a ß-
697   galactosidase (used above) and the resulting glucose was oxidized with a glucose oxidase (Sigma
698   Aldrich) concomitant with the production of H_2O_2 measured by coupling to horseradish peroxidase
699   (Sigma Aldrich) oxidation of 4-aminoantipyrine and 3,5-dichloro-2-hydroxybensensulfonic acid.
700   Reactions were prepared in 96-well microtiter plates to a final volume of 150 μL, containing
701   substrate, lacto-*N*-biosidase, ß-galactosidase (150U mL^{-1}), glucose oxidase (150U mL^{-1}),
702   horseradish peroxidase (150U mL^{-1}), 10 mM 3,5-dichloro-2-hydroxybensensulfonic acid, 1 mM 4-
703   aminoantipyrine in standard assay buffer. Reactions were performed at 37°C and $A_{515\ nM}$ was
704   measured in 5 sec intervals for 30 min. Blanks were prepared by substituting lacto-*N*-biosidase
705   with standard assay buffer in the reaction mixture and a lactose standard (3 μM–500 μM) was
706   used for the quantification.
707   Hydrolysis kinetics of *Rh*Lnb136 (40 nM) and *Er*Lnb136 (10 nM) towards LNT (0.2–5 mM for
708   *Rh*Lnb136 and 0.1–2.5 mM for *Er*Lnb136) were determined as described above. The kinetic
709   parameters K_M and k_{cat}, were calculated by fitting the Michaelis-Menten equation to the initial rate
710   data using OriginPro 2018b (OriginLab). Lacto-*N*-biosidase specific activity of *Ri*Le$^{a/b}$136 (1.2 μM)
711   was measured as described above using 3.5 mM LNT. The specific activity was expressed in
712   units (U) mg^{-1} enzyme, where a unit is defined as the amount of enzyme that releases 1 μmol
713   lactose min^{-1} quantified as above.

714   Specific activities of *Rh*GLnbp112 and *Ri*GLnbp112 towards LNB and GNB were assayed 50 mM
715   sodium phosphate buffer, 150 mM NaCl, 0.005% (v/v) Triton X-100, pH 6.5. Reactions (150 μL)
716   were incubated for 10 min at 37°C with 20 nM enzyme and 2 mM substrate. Aliquots of 15 μL
717   were removed every minute and quenched in 135 μL 0.2 M NaOH. Standards of Gal1P (5
718   mM–0.02 mM) were prepared in 0.2 M NaOH and were used to quantify the concentrations of
719   released Gal1P in the quenched reaction samples. Both, quenched reactions and standards were
720   examined by HPAEC-PAD using a 3 × 250 mm CarboPac PA200 column (Theromofisher) in
721   combination with a 3 × 50 mm CarboPac guard column (Theromofisher) and 10 μL injections.
722   Elution was performed with a flow of 0.350 mL min^{-1} and a mobile phase of 150 mM NaOH
723   and 60 mM sodium acetate. The specific activity was expressed in U mg^{-1} enzyme, where a U is
724   defined as the amount of enzyme that releases 1 μmoL Gal1P min^{-1}. The analysis was performed
725   in technical triplicates.

Enzyme product profiles

727   Enzyme assays were performed at 37°C for 16 h in standard assay buffer or in the phosphate
728   version (instead of MES) for GH112 enzymes. Degradation products were analyzed by thin layer
729   chromatography (TLC) and or Matrix-assisted laser desorption/ionization time of flight mass
730   spectroscopy (MALDI-TOF/MS) as described below.

Thin layer chromatography

The TLC was performed by spotting 2 µL of enzymatic reaction on a silica gel 60 F454 plate (Merck), the separation was carried out in butanol: ethanol: milliQ water (5:3:2) (v/v) as mobile phase and sugars were visualized with 5-methylresorcinol:ethanol:sulfuric acid (2:80:10) (v/v) and heat treatment except for *Ri*Le[a/b]136. The TLC for the latter enzyme was performed in butanol:acetic acid: milliQ (2:1:1)(v/v) and developed with diphenylamine-phosphoric acid reagent[56].

MALDI-TOF/MS

MALDI-TOF/MS analysis of *Ri*Le[a/b]136 was accordind to[57], following permethylation of oligosaccharides[58]. Permethylated sugars were dried, mixed with 2,5-dihydroxybenzoic acid, and spotted onto the MALDI plate. For MALFI-TOF/MS analyses, a Bruker Autoflex III smartbeam in positive ion mode was used. Degradation products of *Rh*Lnb136 and *Er*Lnb136 were analyzed without initial permethylation of oligosaccharides using 2,5-dihydroxybenzoic acid as matrix and an Ultraflex II TOF/TOF (Bruker Daltonics) instrument operated in positive ion linear mode. Peak analysis of mass spectra was performed using Flexanalysis Version 3.3 (Bruker Daltonics).

LC-MS[2] of *O*-glycan derived oligosaccharides

A homogenous preparation of porcine gastric mucin, PGM (Sigma), carrying blood group A, was used in the analysis. A total of 0.1 mg mucin per dot were immobilized by dot blotting onto an immobilon-P PVDF membranes (Immobilon P membranes, 0.45 µm, Millipore, Billerica, MA). *Ri*GH98 was added to one dot to 1.5 µM in 50 µL and incubated for 1 h and 4 h at 37°C. The reaction supernatants which contained released free oligosaccharides, were collected and purified by passage through porous graphitized carbon (PGC) particles (Thermo Scientific) packed on top of a C18 Zip-tip (Millipore). Samples were eluted with 65% (v/v) ACN in 0.5% trifluoro-acetic acid (TFA, v/v), dried, resuspended in 10 µL of milliQ, frozen at -20 °C and stored until further analysis. The residual *O*-linked glycans (on the dot) were released by reductive β-elimination by incubating the dot in 30 µL of 0.5 M $NaBH_4$ in 50 mM NaOH at 50°C for 16 h followed by adding 1.5 µL glacial acetic acid to quench the reaction. The released *O*-glycans were desalted and dried as described before [59]. The purified glycans were resuspended in 10 µL of milliQ and stored at -20°C for further analysis. Released oligosaccharides from glycosphingolipids as a model substrate carrying blood group B (B5-2 and B6-2)[60] were prepared as described above, except for a single incubation time of 2 h.

Purified samples were analyzed by LC-MS/MS using 10 cm × 250 µm I.D. column, packed in house with PGC 5 µm particles. Glycans were eluted using a linear gradient of 0–40% ACN in 10 mM NH_4HCO_3 over 40 min at 10 µl min[-1]. The eluted *O*-glycans were analyzed on a LTQ mass spectrometer (Thermo Scientific) in negative-ion mode with an electrospray voltage of 3.5 kV, capillary voltage of -33.0 V and capillary temperature of 300 °C. Air was used as a sheath gas and mass ranges were defined depending on the specific structure to be analyzed. The data were processed using Xcalibur software (version 2.0.7, Thermo Scientific).

Oligosaccharide binding analysis

773 Binding of LNT, LNB, GNB, H type I triose, Le^a triose and Le^b tetraose to RiLe$^{a/b}$BP was analyzed
774 by surface plasmon resonance (SPR; Biacore T100, GE Healthcare). RiLe$^{a/b}$BP, diluted in 10 mM
775 NaOAc buffer pH 3.75 to 50 µg mL^{-1}, was immobilized on a CM5 chip using a random amine
776 coupling kit (GE Healthcare) to a final chip density of 3214 and 4559 response units (RU). Analysis
777 comprised 90 s for association and 240 s for dissociation phase, respectively, at a flow rate of 30
778 µL min^{-1}. Sensograms were recorded at 25°C in 20 mM sodium phosphate buffer, 150 mM NaCl,
779 0.005% (v/v) P20 (GE Healthcare), pH 6.5. Experiments were performed in duplicates (each
780 consisting of a technical duplicate) in the range of 0.3-50 µM for LNB, 0.78-200 µM for GNB,
781 0.97-250 µM for Le^a, 0.097 µM-100 µM for Le^b and 1.5-250 µM for blood H type I triose. To
782 investigate ligand specify of RiLe$^{a/b}$BP, binding was further tested towards 0.5 mM LNT, LNnT,
783 lactose, blood A triose, 2'FL and 3'FL. Equilibrium dissociation constants (K_D) were calculated by
784 fitting a one binding site model to steady state sensograms, using the Biacore T100 data
785 evaluation software.

786 Binding of LNT, LNB, GNB, LNnT, lactose and 2'FL to RhLNBBP was measured using a Microcal
787 ITC$_{200}$ calorimeter (GE Healthcare). Titrations were performed in duplicates at 25°C with
788 RhLNBBP (0.1 mM) in the sample cell and 1.5 mM ligand in 10 mM sodium phosphate buffer, pH
789 6.5 in the syringe. A first injection of 0.4 µL was followed by 19 injections of 2 µL ligand each,
790 separated by 180 s. Heat of dilution was determined from buffer titrations and corrected data were
791 analyzed using MicroCal Origin software v7.0. To determine binding thermodynamics a non-linear
792 single binding model was fitted to the normalized integrated binding isotherms.

793 **Differential scanning calorimetry (DSC)**

794 The DSC analyses was performed at protein concentrations of 1 mg mL^{-1} in 20 mM sodium
795 phosphate buffer, 150 mM NaCl, pH 6.5, using a Nano DSC (TA instruments). Thermograms
796 were recorded from 10 to 90°C at a scan speed of 1°C min^{-1} using buffer as reference. Baseline
797 corrected data were analyzed using the NanoAnalyze software (TA instruments). DSC analyses
798 were performed in duplicates unless otherwise states.

799 **Crystallization**

800 Crystals of ErLnb136 proteins were grown at 20°C using the sitting-drop vapor diffusion method,
801 by mixing 0.5 µL of a 10 mg mL^{-1} protein solution with an equal volume of a reservoir solution.
802 Native crystals were grown in a 20% (w/v) PEG4000, 0.1 M sodium citrate pH 5.6, and 20%
803 isopropanol reservoir solution. SeMet-labelled crystals were grown using a reservoir solution
804 containing 20% (w/v) PEG6000, 0.1 M Tris-HCl pH 8.5, and 1 M lithium chloride. The crystals
805 were cryoprotected in the reservoir solution supplemented with 20% (v/v) glycerol and 25 mM
806 LNB. The crystals were flash-cooled at 100 K (−173.15°C) in a stream of nitrogen gas. Diffraction
807 data were collected at 100 K on beamlines at SLS X06DA (Swiss Light Source, Swiss) and Photon
808 Factory of the High Energy Accelerator Research Organization (KEK, Tsukuba, Japan). The data
809 were processed using HKL2000[61] and XDS[62]. Initial phase calculation, phase improvement, and
810 automated model building were performed using PHENIX[63]. Manual model rebuilding
811 and refinement was achieved using Coot[64] and REFMAC5[65]. Because the crystal structures of
812 SeMet-labelled and native protein were virtually the same (root mean square deviations for the
813 Cα atoms = 0.14 Å), we used the SeMet-labelled protein structure for the descriptions in the
814 Results and Discussion. Molecular graphics were prepared using PyMOL (Schrödinger, LLC, New
815 York) or UCSF Chimera (University of California, San Francisco)

816 **Bioinformatics**

817 SignalP v.4.1[66], PSORTb v3.0[67], TMHMM v.2.0[68] were used for prediction of signal peptides and
818 transmembrane domains. InterPro[69] and dbCAN2[70] were used to analyse modular organization
819 using default settings for Gram positive bacteria. Redundancy in biological sequence datasets
820 was reduced using the CD-HIT server (sequence identity cut off = 0.95)[71]. Protein sequence
821 alignments were performed using MAFFT (BLOSUM62)[72]. Phylogenetic trees were constructed
822 using the MAFFT server, based on the neighbor-joining algorithm, and with bootstraps performed
823 with 1000 replicates. Phylogenetic trees were visualized and tanglegrams constructed using
824 dendroscope[73]. Coloring of protein structures according to amino acid sequence conservation
825 was accomplished in UCSF Chimera, based on protein multiple (structural based) alignments
826 from the PROMALS3D server[74] and by using the in UCSF Chimera implemented AL2CO
827 algorithm[75]. The MEME suite web server was used for amino acid sequence motif discovery and
828 evaluation[76]. Protein structures were compared using the Dali server
829 (http://ekhidna2.biocenter.helsinki.fi/dali/) (PMID: 27131377) and the molecular interface between
830 ErLnb136$_I$ and ErLnb136$_{II}$ was analyzed (solvent inaccessible interface, Gibbs energy) via the
831 PDBePISA server (https://www.ebi.ac.uk/pdbe/pisa/).

832 The abundance and distribution of HMO utilization genes encoding GH112, GH136$_I$ and GH136$_{II}$
833 in $Roseburia$ were analyzed by a BLAST search of the corresponding DNA reference sequences
834 from $R.$ $intestinalis$ L1-82, $R.$ $hominis$ A2-183 and $R.$ $inulinivorans$ A2-194 against a total of 4599
835 reconstructed $Roseburia$ genomes, binned into 42 Species-level Genome Bins (SGBs) by Pasolli
836 et al.[31]. The variability of the $Roseburia$ core xylanase (GH10) was determined similarly by blasting
837 the DNA reference sequences from $R.$ $intestinalis$ L1-82 (ROSINTL182_06494) against the same
838 dataset.

839 For further analyses, initial blast hits were filtered based on a 70% identity with any of the 5
840 conserved $Roseburia$ reference genomes. Additionally, Roseburia genomes were considered
841 only if they have a hit with GH112 gene. The resulting 1397 genomes were assigned into the
842 respective $Roseburia$ SGBs, base on they assignation of Pasolli et al.[31]. The retrieved genomes
843 were used to analyze the gene landscape around the GH112 gene. The RAST server[77] was used
844 for gene annotation. Based on the annotation and coordinates of the genes, 10 genes upstream
845 and downstream the GH112 were selected for gene landscapes analysis.

846 <u>Quantification and Statistical Analysis:</u>

847 Statistical significant differences were determined using unpaired two-tailed Student's t-test.
848 Statistical parameters, including values of n and p-values, are reported or indicated in the figures,
849 figure legends and the result section. The data are expressed as arithmetic means with standard
850 deviations (SD), unless otherwise indicated.

851 <u>Data and Code availability:</u>

852 The mass spectrometry proteomics data have been deposited to the ProteomeXchange
853 Consortium via the PRIDE partner repository with the dataset identifier PXD015045. The
854 accession numbers for the atomic coordinates reported in this paper are PDB: 6KQS (Se-Met)
855 and 6KQT (native), see also Table S6. Mucin glycomis MS/MS data are summarized in Table
856 S9 and raw data files are available upon request.

857

858   <u>Material and Resource availability:</u>

859   Requests for resources and material should be addressed to Maher Abou Hachem
860   (maha@bio.dtu.dk)

861

862

863

864

References

1. Clemente, J. C., Ursell, L. K., Parfrey, L. W. & Knight, R. The impact of the gut microbiota on human health: An integrative view. *Cell* **148**, 1258–1270 (2012).

2. Sonnenburg, J. L. & Bäckhed, F. Diet-microbiota interactions as moderators of human metabolism. *Nature* **535**, 56–64 (2016).

3. Belkaid, Y. & Hand, T. W. Role of the microbiota in immunity and inflammation. *Cell* **157**, 121–41 (2014).

4. Bokulich, N. A. *et al.* Antibiotics, birth mode, and diet shape microbiome maturation during early life. *Sci. Transl. Med.* **8**, 343ra82 (2016).

5. Yassour, M. *et al.* Strain-Level Analysis of Mother-to-Child Bacterial Transmission during the First Few Months of Life. *Cell Host Microbe* **24**, 146-154.e4 (2018).

6. Bäckhed, F. *et al.* Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host Microbe* **17**, 690–703 (2015).

7. Stewart, C. J. *et al.* Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature* **562**, 583–588 (2018).

8. Yatsunenko, T. *et al.* Human gut microbiome viewed across age and geography. *Nature* **486**, 222–227 (2012).

9. Robertson, R. C., Manges, A. R., Finlay, B. B. & Prendergast, A. J. The Human Microbiome and Child Growth – First 1000 Days and Beyond. *Trends in Microbiology* **27**, 131–147 (2019).

10. Cox, L. M. *et al.* Altering the intestinal microbiota during a critical developmental window has lasting metabolic consequences. *Cell* **158**, 705–721 (2014).

11. Cox, L. M. & Blaser, M. J. Antibiotics in early life and obesity. *Nature Reviews Endocrinology* **11**, 182–190 (2015).

12. Tamburini, S., Shen, N., Wu, H. C. & Clemente, J. C. The microbiome in early life: implications for health outcomes. *Nat. Med.* **22**, 713–22 (2016).

13. Marchesi, J. R. *et al.* The gut microbiota and host health: a new clinical frontier. *Gut* **65**, 330–9 (2016).

14. Vogt, N. M. *et al.* Gut microbiome alterations in Alzheimer's disease. *Sci. Rep.* **7**, 13537 (2017).

15. Koropatkin, N. M., Cameron, E. A. & Martens, E. C. How glycan metabolism shapes the human gut microbiota. *Nat. Rev. Microbiol.* **10**, 323–35 (2012).

16. Gotoh, A., Ojima, M. N. & Katayama, T. Minority species influences microbiota formation: the role of Bifidobacterium with extracellular glycosidases in bifidus flora formation in breastfed infant guts. *Microb. Biotechnol.* **12**, 259–264 (2019).

17. Sakanaka, M. *et al.* Evolutionary adaptation in fucosyllactose uptake systems supports bifidobacteria-infant symbiosis. *Sci. Adv.* **5**, eaaw7696 (2019).

18. Norin, E., Midtvedt, T. & Björkstén, B. Development of faecal short-chain fatty acid

903   pattern during the first year of life in Estonian and Swedish infants. *Microb. Ecol. Health*
904   *Dis.* **16**, 8–12 (2004).

905 19. Furusawa, Y. *et al.* Commensal microbe-derived butyrate induces the differentiation of
906   colonic regulatory T cells. *Nature* **504**, 446–450 (2013).

907 20. Bultman, S. J. Bacterial butyrate prevents atherosclerosis. *Nat. Microbiol.* **3**, 1332–1333
908   (2018).

909 21. Canani, R. B. *et al.* Potential beneficial effects of butyrate in intestinal and extraintestinal
910   diseases. *World J. Gastroenterol.* **17**, 1519–28 (2011).

911 22. Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473**, 174–180
912   (2011).

913 23. Louis, P. & Flint, H. J. Diversity, metabolism and microbial ecology of butyrate-producing
914   bacteria from the human large intestine. *FEMS Microbiol. Lett.* **294**, 1–8 (2009).

915 24. Karlsson, F. H. *et al.* Symptomatic atherosclerosis is associated with an altered gut
916   metagenome. *Nat. Commun.* **3**, 1245 (2012).

917 25. Le Chatelier, E. *et al.* Richness of human gut microbiome correlates with metabolic
918   markers. *Nature* **500**, 541–546 (2013).

919 26. Wang, J. *et al.* A metagenome-wide association study of gut microbiota in type 2
920   diabetes. *Nature* **490**, 55–60 (2012).

921 27. Vital, M., Karch, A. & Pieper, D. H. Colonic Butyrate-Producing Communities in Humans:
922   an Overview Using Omics Data. *mSystems* **2**, (2017).

923 28. Yamada, C. *et al.* Molecular Insight into Evolution of Symbiosis between Breast-Fed
924   Infants and a Member of the Human Gut Microbiome Bifidobacterium longum. *Cell Chem.*
925   *Biol.* **24**, 515-524.e5 (2017).

926 29. Sakurama, H. *et al.* Lacto-N-biosidase encoded by a novel gene of *Bifidobacterium*
927   *longum* subspecies *longum* shows unique substrate specificity and requires a designated
928   chaperone for its active expression. *J. Biol. Chem.* **288**, 25194–206 (2013).

929 30. Derrien, M., Vaughan, E. E., Plugge, C. M. & de Vos, W. M. *Akkermansia municiphila*
930   gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *Int. J. Syst. Evol.*
931   *Microbiol.* **54**, 1469–1476 (2004).

932 31. Pasolli, E. *et al.* Extensive Unexplored Human Microbiome Diversity Revealed by Over
933   150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell*
934   **176**, 649-662.e20 (2019).

935 32. Leth, M. L. *et al.* Differential bacterial capture and transport preferences facilitate co-
936   growth on dietary xylan in the human gut. *Nat. Microbiol.* **3**, 570–580 (2018).

937 33. Wang, T. *et al.* Structural segregation of gut microbiota between colorectal cancer
938   patients and healthy volunteers. *ISME J.* **6**, 320–329 (2012).

939 34. Theilmann, M. C., Fredslund, F., Svensson, B., Lo Leggio, L. & Abou Hachem, M.
940   Substrate preference of an ABC importer corresponds to selective growth on β-(1,6)-
941   galactosides in *Bifidobacterium animalis* subsp. *lactis*. *J. Biol. Chem.* jbc.RA119.008843
942   (2019). doi:10.1074/jbc.ra119.008843

943   35.   Ejby, M. *et al.* Two binding proteins of the ABC transporter that confers growth of
944   *Bifidobacterium animalis* subsp. *lactis* ATCC27673 on β-mannan possess distinct manno-
945   oligosaccharide-binding profiles. *Mol. Microbiol.* **112**, 114–130 (2019).

946   36.   Kostic, A. D. *et al.* The dynamics of the human infant gut microbiome in development and
947   in progression toward type 1 diabetes. *Cell Host Microbe* **17**, 260–273 (2015).

948   37.   Jost, T., Lacroix, C., Braegger, C. & Chassard, C. Assessment of bacterial diversity in
949   breast milk using culture-dependent and culture-independent approaches. *Br. J. Nutr.*
950   **110**, 1253–1262 (2013).

951   38.   Collado, M. C., Delgado, S., Maldonado, A. & Rodríguez, J. M. Assessment of the
952   bacterial diversity of breast milk of healthy women by quantitative real-time PCR. *Lett.*
953   *Appl. Microbiol.* **48**, 523–528 (2009).

954   39.   Jost, T., Lacroix, C., Braegger, C. P., Rochat, F. & Chassard, C. Vertical mother-neonate
955   transfer of maternal gut bacteria via breastfeeding. *Environ. Microbiol.* **16**, 2891–2904
956   (2014).

957   40.   Duncan, S. H., Hold, G. L., Barcenilla, A., Stewart, C. S. & Flint, H. J. Roseburia
958   intestinalis sp. nov., a novel saccharolytic, butyrate-producing bacterium from human
959   faeces. *Int. J. Syst. Evol. Microbiol.* **52**, 1615–1620 (2002).

960   41.   La Rosa, S. L. *et al.* The human gut Firmicute *Roseburia intestinalis* is a primary
961   degrader of dietary β-mannans. *Nat. Commun.* **10**, 905 (2019).

962   42.   Scott, K. P., Martin, J. C., Duncan, S. H. & Flint, H. J. Prebiotic stimulation of human
963   colonic butyrate-producing bacteria and bifidobacteria, *in vitro*. *FEMS Microbiol. Ecol.* **87**,
964   30–40 (2014).

965   43.   Rossez, Y. *et al.* Almost all human gastric mucin *O*-glycans harbor blood group A, B or H
966   antigens and are potential binding sites for *Helicobacter pylori*. *Glycobiology* **22**, 1193–
967   1206 (2012).

968   44.   Holgersson, J., Jovall, P. Å. & Breimer, M. E. Glycosphingolipids of human large
969   intestine: Detailed structural characterization with special reference to blood group
970   compounds and bacterial receptor structures. *J. Biochem.* **110**, 120–131 (1991).

971   45.   Anand, S., Kaur, H. & Mande, S. S. Comparative in silico analysis of butyrate production
972   pathways in gut commensals and pathogens. *Front. Microbiol.* **7**, 1945 (2016).

973   46.   Van Den Abbeele, P. *et al.* Butyrate-producing Clostridium cluster XIVa species
974   specifically colonize mucins in an in vitro gut model. *ISME J.* **7**, 949–961 (2013).

975   47.   Ward, R. E. Isolation of Milk Oligosaccharides using Solid-Phase Extraction. *Methodology*
976   **2**, 9–15 (2009).

977   48.   Robinson, R. C., Colet, E., Tian, T., Poulsen, N. A. & Barile, D. An improved method for
978   the purification of milk oligosaccharides by graphitised carbon-solid phase extraction. *Int.*
979   *dairy J.* **80**, 62–68 (2018).

980   49.   Miller, R. S. & Hoskins, L. C. Mucin degradation in human colon ecosystems: Fecal
981   population densities of mucin-degrading bacteria estimated by a "most probable number"
982   method. *Gastroenterology* **81**, 759–765 (1981).

983   50.   Skoog, E. C. *et al.* Human gastric mucins differently regulate *Helicobacter pylori*

984 proliferation, gene expression and interactions with host cells. *PLoS One* **7**, e36378
985 (2012).

986 51. Dümmler, A., Lawrence, A.-M. & de Marco, A. Simplified screening for the detection of
987 soluble fusion constructs expressed in *E. coli* using a modular set of vectors. *Microb. Cell*
988 *Fact.* **4**, 34 (2005).

989 52. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated
990 proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat.*
991 *Methods* **11**, 319–324 (2014).

992 53. Ejby, M. *et al.* An atp binding cassette transporter mediates the uptake of α-(1,6)-linked
993 dietary oligosaccharides in bifidobacterium and correlates with competitive growth on
994 these substrates. *J. Biol. Chem.* **291**, 20220–20231 (2016).

995 54. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-
996 fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**,
997 1896–1906 (2007).

998 55. Gurdeep Singh, R. *et al.* Unipept 4.0: Functional Analysis of Metaproteome Data. *J.*
999 *Proteome Res.* **18**, 606–615 (2019).

1000 56. Anderson, K., Li, S.-C. & Li, Y.-T. Diphenylamine–Aniline–Phosphoric Acid Reagent, a
1001 Versatile Spray Reagent for Revealing Glycoconjugates on Thin-Layer Chromatography
1002 Plates. *Anal. Biochem.* **287**, 337–339 (2000).

1003 57. Sugiyama, Y. *et al.* Introduction of H-antigens into oligosaccharides and sugar chains of
1004 glycoproteins using highly efficient 1,2-α-L-fucosynthase. *Glycobiology* **26**, 1235–1247
1005 (2016).

1006 58. Anumula, K. R. & Taylor, P. B. A comprehensive procedure for preparation of partially
1007 methylated alditol acetates from glycoprotein carbohydrates. *Anal. Biochem.* **203**, 101–
1008 108 (1992).

1009 59. Karlsson, N. G., Schulz, B. L. & Packer, N. H. Structural determination of neutral *O*-linked
1010 oligosaccharide alditols by negative ion LC-electrospray-MSn. *J. Am. Soc. Mass*
1011 *Spectrom.* **15**, 659–672 (2004).

1012 60. Ångström, J. *et al.* Novel carbohydrate binding site recognizing blood group A and B
1013 determinants in a hybrid of cholera toxin and *Escherichia coli* heat-labile enterotoxin B-
1014 subunits. *J. Biol. Chem.* **275**, 3231–3238 (2000).

1015 61. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation
1016 mode. *Methods Enzymol.* **276**, 307–326 (1997).

1017 62. Kabsch, W. XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 125–132 (2010).

1018 63. Adams, P. D. *et al.* PHENIX: Building new software for automated crystallographic
1019 structure determination. *Acta Crystallogr. Sect. D Biol. Crystallog* **58**, 1948–1954
1020 (International Union of Crystallography, 2002).

1021 64. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot.
1022 *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).

1023 65. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures
1024 by the maximum-likelihood method. *Acta Crystallographica Section D: Biological*

1025   *Crystallography* **53**, 240–255 (1997).

1026   66.   Nielsen, H. Predicting secretory proteins with signaIP. in *Methods in Molecular Biology*
1027   **1611**, 59–73 (Humana Press, New York, NY, 2017).

1028   67.   Yu, N. Y. *et al.* PSORTb 3.0: improved protein subcellular localization prediction with
1029   refined localization subcategories and predictive capabilities for all prokaryotes.
1030   *Bioinformatics* **26**, 1608–1615 (2010).

1031   68.   Krogh, A., Larsson, B., Von Heijne, G. & Sonnhammer, E. L. . Predicting transmembrane
1032   protein topology with a hidden Markov model: Application to complete genomes. *J. Mol.*
1033   *Biol.* **305**, 567–580 (2001).

1034   69.   Mitchell, A. L. *et al.* InterPro in 2019: Improving coverage, classification and access to
1035   protein sequence annotations. *Nucleic Acids Res.* **47**, D351–D360 (2019).

1036   70.   Zhang, H. *et al.* DbCAN2: A meta server for automated carbohydrate-active enzyme
1037   annotation. *Nucleic Acids Res.* **46**, W95–W101 (2018).

1038   71.   Huang, Y., Niu, B., Gao, Y., Fu, L. & Li, W. CD-HIT Suite: A web server for clustering and
1039   comparing biological sequences. *Bioinformatics* **26**, 680–682 (2010).

1040   72.   Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: multiple sequence
1041   alignment, interactive sequence choice and visualization. *Brief. Bioinform.* (2017).
1042   doi:10.1093/bib/bbx108

1043   73.   Huson, D. H. *et al.* Dendroscope: An interactive viewer for large phylogenetic trees. *BMC*
1044   *Bioinformatics* **8**, 460 (2007).

1045   74.   Pei, J., Kim, B. H. & Grishin, N. V. PROMALS3D: A tool for multiple protein sequence and
1046   structure alignments. *Nucleic Acids Res.* **36**, 2295–2300 (2008).

1047   75.   Pei, J. & Grishin, N. V. AL2CO: Calculation of positional conservation in a protein
1048   sequence alignment. *Bioinformatics* **17**, 700–712 (2001).

1049   76.   Bailey, T. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res*
1050   **37**, (2009).

1051   77.   Aziz, R. K. *et al.* The RAST Server: rapid annotations using subsystems technology. *BMC*
1052   *Genomics* **9**, 75 (2008).

1053

1054