Butyrate producing colonic Clostridiales metabolise human milk oligosaccharides and cross feed on mucin via conserved pathways

Pichler, Michael Jakob; Yamada, Chihaya; Shuoker, Bashar; Alvarez-Silva, Camila; Gotoh, Aina; Leth, Maria Louise; Schoof, Erwin; Katoh, Toshihiko; Sakanaka, Mikiyasu; Katayama, Takane; Jin, Chunsheng; Karlsson, Niclas G.; Arumugam, Manimozhiyan; Fushinobu, Shinya; Abou Hachem, Maher

Published in:
Nature Communications

DOI:
10.1038/s41467-020-17075-x

Publication date:
2020

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
Pichler, M. J., Yamada, C., Shuoker, B., Alvarez-Silva, C., Gotoh, A., Leth, M. L., Schoof, E., Katoh, T., Sakanaka, M., Katayama, T., Jin, C., Karlsson, N. G., Arumugam, M., Fushinobu, S., & Abou Hachem, M. (2020). Butyrate producing colonic Clostridiales metabolise human milk oligosaccharides and cross feed on mucin via conserved pathways. Nature Communications, 11, [3285]. https://doi.org/10.1038/s41467-020-17075-x
Butyrate producing colonic Clostridiales metabolise human milk oligosaccharides and cross feed on mucin via conserved pathways

Michael Jakob Pichler1,8, Chihaya Yamada2,8, Bashar Shuoker1,3, Camila Alvarez-Silva4, Aina Gotoh5, Maria Louise Leth1, Erwin Schoof1, Toshihiko Katoh5, Mikiyasu Sakanaka6, Takane Katayama5,6, Chunsheng Jin7, Niclas G. Karlsson7, Manimozhiyan Arumugam4, Shinya Fushinobu2 & Maher Abou Hachem1

The early life human gut microbiota exerts life-long health effects on the host, but the mechanisms underpinning its assembly remain elusive. Particularly, the early colonization of Clostridiales from the Roseburia-Eubacterium group, associated with protection from colorectal cancer, immune- and metabolic disorders is enigmatic. Here, we describe catabolic pathways that support the growth of Roseburia and Eubacterium members on distinct human milk oligosaccharides (HMOs). The HMO pathways, which include enzymes with a previously unknown structural fold and specificity, were upregulated together with additional glycan-utilization loci during growth on selected HMOs and in co-cultures with Akkermansia muciniphila on mucin, suggesting an additional role in enabling cross-feeding and access to mucin O-glycans. Analyses of 4599 Roseburia genomes underscored the preponderance and diversity of the HMO utilization loci within the genus. The catabolism of HMOs by butyrate-producing Clostridiales may contribute to the competitiveness of this group during the weaning-triggered maturation of the microbiota.
The human gut microbiota (HGM) is a key determinant of health\(^1\)–\(^3\). Orthogonal transfer from the mother contributes markedly to the establishment of this community shortly after birth\(^4\)–\(^9\). The HGM develops dynamically during infancy until a resilient adult-like community is formed after 2–3 years of life\(^10\)–\(^12\). The early life microbiota plays a role in the maturation of the host’s endocrine, metabolic and immune systems\(^9\), and the composition of this consortium is associated with lifelong health effects\(^10\)–\(^12\). Therefore, understanding the factors that define the HGM structure during infancy is critical for minimizing the risk for a range of metabolic, inflammatory and neurodegenerative disorders, all associated to specific HGM signatures\(^13\)–\(^14\).

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

Butyrate exerts immune-modulatory activities\(^19\) and is associated with a lowered risk of colon cancer, atherosclerosis, and enteric colitis\(^20\)–\(^21\). The production of butyrate is largely ascribed to *Firmicutes* \(^2\) that de novo ferment human milk oligosaccharides (HMOs), which potentially may confer an advantage during the maturation of the infant HGM during weaning. Genomic analyses of butyrate producers from Lachnospiraceae identified distant homologs of the recently discovered glycoside hydrolase family 136 (GH136) (GH136) in the Carbohydrate Active enZyme (CAZy) 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 3.2.1.146; Supplementary Table 1). The activity of the bifidobacterial LnbX was dependent on the co-expression of an adjacent gene, proposed to encode a molecular chaperone (LnbY). The GH136 orthologues from *Roseburia* and *Eubacterium* are organized, unlike the counterpart from *Bifidobacterium*, on a locus harbouring additional CAZy genes (Supplementary Fig. 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 \((p < 4 \times 10^{-4})\) and *Roseburia inulinivorans* DSM 16841 \((p < 1.3 \times 10^{-4})\) after 24 h on media with HMOs from mother’s milk, but the growth of *R. inulinivorans* was more efficient \((\mu_{\text{max}} = 0.30 \pm 0.01 \text{ h}^{-1})\) (Fig. 1a,b). Next, we carried out growth on building blocks from HMOs and related oligomers from O-glycoconjugates (Fig. 1a–d). None of the strains grew on the top abundant fucosyl-lactose (FL) HMOs, despite good growth on lactose (Fig. 1d). *Roseburia* strains failed to grow on sialyl lactose (SL) (Fig. 1d), consistent with the lack of encoded sialidases. *R. hominis* grew efficiently on the HMO LNT \((\mu_{\text{max}} = 0.22 \pm 0.02 \text{ h}^{-1})\), its LNB unit \((\mu_{\text{max}} = 0.16 \pm 0.01 \text{ h}^{-1})\) and the mucin-derived galacto-N-biose (GNB) \((\mu_{\text{max}} = 0.21 \pm 0.02 \text{ h}^{-1})\) (Fig. 1a). Growth on LNT was also shared by the taxonomically related *Eubacterium ramosum* DSM 15684 from Eubacteriaceae. By contrast, *R. inulinivorans* grew well only on LNB and GNB, but not LNT (Fig. 1c). *R. inulinivorans* was further distinguished by growth on sialic acid (Neu5Ac), abundant in HMOs and glycoconjugates (Fig. 1d).

Typically, bacteria repress pathways for less preferred substrates in the presence of a favourable carbon source. Xylotetraose from the abundant dietary plant fibre xylan has been shown to be a preferred growth substrate and uptake ligand of the xylo-oligosaccharide importer conserved in *Roseburia*\(^29\). During weaning, the HGM of infants is likely to be exposed to both HMOs and dietary plant fibres, e.g. xylan from cereals and fruits. We tested the growth of *R. hominis* in the presence of equimolar concentrations of LNT and the similarly sized xylotetraose to evaluate the utilization hierarchy of the HMO versus the plant fibre. Strikingly, monophasic growth was observed consistent with the simultaneous uptake of both tetraoses from culture supernatant (Supplementary Fig. 2a–c).

To unravel the basis of growth on HMOs with focus on the *Roseburia* genus, we analysed the 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 \((\log_2 \text{ fold change} > 2)\). These differential proteomes were dominated by carbohydrate metabolism proteins, especially products of two loci encoding the resilience of key butyrate-producing Clostridiales by mediating the catabolism of distinct HMOs and host O-glycans.

**Results**

**HMO loci in *Roseburia* and *Eubacterium***. Our aim was to investigate the HMO utilization potential in butyrate producing Clostridiales, which potentially may confer an advantage during the maturation of the infant HGM during weaning. Genomic analyses of butyrate producers from Lachnospiraceae identified distant homologs of the recently discovered glycoside hydrolase family 136 (GH136) (GH136) in the Carbohydrate Active enZyme (CAZy) 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 3.2.1.146; Supplementary Table 1). The activity of the bifidobacterial LnbX was dependent on the co-expression of an adjacent gene, proposed to encode a molecular chaperone (LnbY). The GH136 orthologues from *Roseburia* and *Eubacterium* are organized, unlike the counterpart from *Bifidobacterium*, on a locus harbouring additional CAZy genes (Supplementary Fig. 1).

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

The evolution of uptake and enzymatic systems that support competitive growth of *Bifidobacterium* spp. on HMOs\(^17\) reflects a successful adaptation to the intestines of breast fed infants. We hypothesize that other taxonomic groups, which possess metabolic capabilities that target HMOs, may have an early advantage in the colonization of the infant gut during infancy.

The early emergence of *Roseburia-Eubacterium* in the human gut offers a suitable model group to evaluate this hypothesis. Here, we perform genomic analyses that are suggestive of the presence of putative HMO utilization loci in *Roseburia* and *Eubacterium* strains. Growth on selected HMOs or a complex mixture from mother’s milk combined with differential proteomics reveal the high upregulation of the protein apparatus encoded by these loci, consistent with their role in mediating HMO utilization.

Further, we characterize enzymes and transport proteins encoded by the HMO loci to elucidate the molecular details of HMO capture and degradation by this protein apparatus. These analyses unveil an enzymatic activity and a structural fold, which have not been previously reported. The HMO catabolic pathways are upregulated during growth with the model mucin degrader *Akkermansia muciniphila*, suggesting these pathways may support cross-feeding on mucin oligosaccharides made accessible by *A. muciniphila*. Analyses of the metagenome of *Roseburia* show a striking conservation and wide occurrence of the HMO utilization pathways across the genus, underscoring their importance for adaptation to the human gut. This study provides insight into pathways that may confer a competitive advantage in the early colonization and the resilience of key butyrate-producing Clostridiales by mediating the catabolism of distinct HMOs and host O-glycans.
Glycoside hydrolase 112 (transporter permease protein (PP); hypothetical proteins (HP); Glycoside hydrolase 136 (gov/glycans/snfg.html). Source data are provided as a Source data upregulated gene products is shown. Glycan structures presentation according to Symbol Nomenclature for Glycans (SNFG) (https://www.ncbi.nlm.nih.gov/glycans). The proteomic analyses (inulinivorans of R. inulinivorans) were performed in independent biological triplicates. The growth data are presented as mean values with the error bars representing the standard deviations (SD) for a-c. e HMO and mucin oligomeric growth substrates in a-d. The HMO utilization loci in R. hominis (f) and R. inulinivorans (g) identified from proteomic analyses of cells growing on LNT and HMOs from mother’s milk, respectively, relative to glucose. Genes are denoted by their protein products: transcriptional regulator (Trans. R.); ABC transporter solute binding protein (RhlNBBP) and RlLe136BP (g); ABC transporter permease protein (PP); hypothetical proteins (HP); Glycoside hydrolase 136 (Rlnb136, Rlnb136c, and RlLe136c, RlLe136c (g)); Glycoside hydrolase 112 (RglNbp112) and Glycoside hydrolase 29 (Rfuc29 (g)); Glycoside hydrolase 95 (Rfuc95 (g)) and histidine kinase sensory protein (His. K.). The proteomic analyses (f-g) were in biological triplicates and the log2-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). Source data are provided as a Source data file labelled with the corresponding figure number and panel definition.

Fig. 1 Growth of Roseburia and Eubacterium spp. on HMOs and upregulation of HMOs utilization loci in Roseburia. Growth curves of R. hominis (a) and R. inulinivorans (b) on glucose, LNT, GNB, LNB, and/or purified HMOs from mother’s milk compared to no-carbon source controls over 24 h. e Growth levels of R. inulinivorans on LNT, LNB, GNB and of E. ramulus on LNT within 24 h including glucose and a no-carbon source controls. d Growth of R. hominis, R. inulinivorans and E. ramulus on lactose, 2′FL, 3′FL, 3′SL and 6′SL as well as on monosaccharides from HMOs and mucin after 24 h including a non-carbon source control. Growth analyses (a-d) on media supplemented with 0.5 % (w/v) carbohydrates (for R. inulinivorans on 1% (w/v) and 4% (w/v) purified HMOs from mothers milk) were performed in independent biological triplicates. The growth data are presented as mean values with the error bars representing the standard deviations (SD) for a-c. e HMO and mucin oligomeric growth substrates in a-d. The HMO utilization loci in R. hominis (f) and R. inulinivorans (g) identified from proteomic analyses of cells growing on LNT and HMOs from mother’s milk, respectively, relative to glucose. Genes are denoted by their protein products: transcriptional regulator (Trans. R.); ABC transporter solute binding protein (RhlNBBP) and RlLe136BP (g); ABC transporter permease protein (PP); hypothetical proteins (HP); Glycoside hydrolase 136 (Rlnb136, Rlnb136c, and RlLe136c, RlLe136c (g)); Glycoside hydrolase 112 (RglNbp112) and Glycoside hydrolase 29 (Rfuc29 (g)); Glycoside hydrolase 95 (Rfuc95 (g)) and histidine kinase sensory protein (His. K.). The proteomic analyses (f-g) were in biological triplicates and the log2-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). Source data are provided as a Source data file labelled with the corresponding figure number and panel definition.

Diverse GH136 enzymes mediate initial HMO degradation. The homologs RhlN136 (LbY in B. longum) and RhlN136 (LbnX that harbours the catalytic residues in B. longum) were highly co-upregulated in the LNT proteome of R. hominis (Fig. 1f). Both, RhlN136 and RhlN136 lack a predicted
transmembrane domain and signal peptide in contrast to the *B. longum* counterparts (Supplementary Fig. 4a), indicative of the intracellular degradation of LNT in *R. hominis*. Only co-expression and co-purification of *Rh*Lnb136I and *Rh*Lnb136II resulted in an active lacto-N-biosidase (henceforth *Rh*Lnb136) (Fig. 2b, Supplementary Table 4). These findings and the observed co-upregulation, suggested that a heterodimer (or oligomer) of *Rh*Lnb136I and *Rh*Lnb136II assembles the catalytically active *Rh*Lnb136. Next, we demonstrated phosphorolysis of LNB and GNB to α-D-galactose-1-phosphate and the corresponding N-acetylhexosamines GlcNAc and GalNAc, respectively (Supplementary Fig. 5f), by the GH112 GNB/LNB phosphorylase (*Rh*GLnbp112) located in the same locus (Fig. 1f and Supplementary Fig. 1). This enzyme has comparable specific activities for LNB and GNB (Supplementary Table 5) consistent with the growth on these disaccharides. The functional lacto-N-biosidase and GNB/LNB phosphorylase further support the HMO catalysis role of the locus.

**Fig. 2 Specificities of GH136 enzymes that mediate the HMO degradation.** a Activity of *Ri*Lea/b136 on fucosylated HMOs. b Activity of *Rh*Lnb136 on LNT. c Activity of *Er*Lnb136 on LNT. a-c The hydrolysates were analysed by MALDI-ToF MS without (b, c) or with a permethylation. a Masses of methylated sugars are in parentheses and the ion peaks correspond to the Na⁺ adducts of the methylated sugars. a-c Relative intensity (percentage intensity) is shown. The MALDI-ToF MS analyses (a-c) were performed from independent triplicates (one analysis from each biological enzymatic reaction replicate) and all analyses yielded similar results.
RlGH136, the GH136 homolog from the HMO-upregulated locus in R. inulinivorans was predicted to be extracellular, with a signal peptide in RlGH136\textsubscript{II} that also possesses two C-terminal putative carbohydrate binding modules (Supplementary Fig. 8a) and a predicted N-terminal transmembrane domain in RlGH136\textsubscript{I}. Co-expression of RlGH136\textsubscript{I} and RlGH136\textsubscript{II}, lacking the transmembrane domain and signal peptide respectively, resulted in an active enzyme with an unprecedented specificity. This enzyme (RlLe\textsubscript{ab}136) released Lewis a triose or Lewis b tetraose from fucosylated HMOs including lacto-N-fucopentaose II (LNFP II), lacto-N-difucohexaose I (LNDFH I) and lacto-N-difucohexaose II (LNDFH II) (Fig. 2a and Supplementary Fig. 5a). To our knowledge, cleavage of the bond at the reducing end of a fucosylated GlcNAc has not been reported to date. Next, we characterized the additional CAZymes encoded by the locus, all lacking a signal peptide or transmembrane domain suggestive of their intracellular localization. We showed that the concerted action of RiFuc29 and RiFuc95 that act on α-(1→4) and α-(1→2)-linked fucosyl, respectively mediates the complete defucosylation of putative products of RlGH136, Le\textsubscript{b} tetraose, Le\textsubscript{a} triose and H triose type I (Supplementary Fig. 5b–d). Initial defucosylation by RiFuc29 is required for releasing the 1→2 linked fucosyl in Le\textsubscript{b} tetraose by RiFuc95. Finally, we showed that the GH112 from R. inulinivorans (RlGlNb1p112) phosphorolyzes LNB and GNB equally efficiently (Supplementary Fig. 5e, Supplementary Table 5).

A domain with a new fold is required for GH136 activity. To discern the molecular architecture of GH136 enzymes and explain the requirement of the two subunits for activity, we endeavoured to crystallize both RhLnb136 and RiLe\textsubscript{ab}136, without success. Hence, we turned our attention to the taxonomically related E. ramulus, which has a GH136 locus similar to the one in R. hominis, except for a substitution of the GLNBP from a GH112 phosphorylase with a GH42-β-galactosidase gene (Supplementary Fig. S1) that may confer hydrolysis of the LNB/GNB units\textsuperscript{30}. The loss of the fucosyl unit at the terminal reducing group H antigen triose type I (H triose type I) relative to Leb\textsubscript{b} tetraose. The specificity of RiLe\textsubscript{ab}136 is highlighted by the lack of activity for lacto-N-neotetrose (LNNT), blood group A antigen triose (A triose), lactose and 2′-fucosyllactose (2′-FL). These findings show that the products of RiLe\textsubscript{ab}136 are the preferred ligands for RiLe\textsubscript{ab}136BP, consistent with the extracellular degradation of fucosylated pentaose and hexaose HMOs and uptake of their products by the ABC transporter. An equimolar mixture of Le\textsubscript{b}, Le\textsubscript{a} and H-triose type I oligomers promoted the growth of R. inulinivorans to a similar final OD\textsubscript{600} as glucose (Fig. 1b, c and Fig. 4a). The uptake profiles of these ligands reflected the preference of RiLe\textsubscript{ab}136BP, consistent with uptake by the associated transporter (Fig. 4). This was also in accord with the utilization of larger fucosylated HMO structures observed during growth on purified HMOs from mother’s milk (Supplementary Fig. 2a–c). Notably, no uptake of LNT was observed, which is in excellent agreement with the poor growth (Fig. 1c) and with the lack of detectable binding to LNT by RiLe\textsubscript{ab}136BP (Fig. 4a).

These results established the capture of specific HMOs and related ligands by the above SBPs and the differentiation of their specificities, e.g. preference of RiLe\textsubscript{ab}136BP to fucosylated ligands at the terminal reducing GlcNAc.

Capture and uptake of HMOs by Roseburia. 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 (Fig. 4, Supplementary Tables 2 and 3, Supplementary Fig. 7). The R. hominis SBP (LNBP-binding protein, RiLNBPB) shows preference to LNB followed by GNB and LNT, suggestive of the uptake and intracellular degradation of these ligands by RhGlNb1p112 and RhLnb136 as described above. By contrast, fucosyl-decorated Lewis b (Le\textsubscript{b}) tetraose and Lewis a (Le\textsubscript{a}) triose were the preferred ligands of the Le\textsubscript{b} binding protein (RiLe\textsubscript{ab}136BP) from R. inulinivorans, followed by LNB and GNB, whereas no binding to LNT was detected (Fig. 4, Supplementary Table 3). The loss of the fucosyl unit at the terminal reducing GlcNAc reduced the affinity of RiLe\textsubscript{ab}136BP about 5-fold for blood group H antigen triose type I (H triose type I) relative to Le\textsubscript{b} tetraose. The specificity of RiLe\textsubscript{ab}136BP is highlighted by the lack of activity for lacto-N-neotetrose (LNNT), blood group A antigen triose (A triose), lactose and 2′-fucosyllactose (2′-FL). These findings show that the products of RiLe\textsubscript{ab}136 are the preferred ligands for RiLe\textsubscript{ab}136BP, consistent with the extracellular degradation of fucosylated pentaose and hexaose HMOs and uptake of their products by the ABC transporter. An equimolar mixture of Le\textsubscript{b}, Le\textsubscript{a} and H-triose type I oligomers promoted the growth of R. inulinivorans to a similar final OD\textsubscript{600} as glucose (Fig. 1b, c and Fig. 4a). The uptake profiles of these ligands reflected the preference of RiLe\textsubscript{ab}136BP, consistent with uptake by the associated transporter (Fig. 4). This was also in accord with the utilization of larger fucosylated HMO structures observed during growth on purified HMOs from mother’s milk (Supplementary Fig. 2a–c). Notably, no uptake of LNT was observed, which is in excellent agreement with the poor growth (Fig. 1c) and with the lack of detectable binding to LNT by RiLe\textsubscript{ab}136BP (Fig. 4a).

These results established the capture of specific HMOs and related ligands by the above SBPs and the differentiation of their specificities, e.g. preference of RiLe\textsubscript{ab}136BP to fucosylated ligands at the terminal reducing GlcNAc.

Roseburia cross feeding on mucin. HMOs and O-glycans from glycolipids and glyco-proteins including mucin share structural motifs. The high affinity of the SBPs from Roseburia for GNB from mucin suggested possible foraging of this substrate (and/or oligomers from glycoconjugates) and thereby a metabolic interplay of Roseburia with mucolytic HG11 members. To evaluate possible mechanisms of cross-feeding we compared Roseburia
growth on mucin with and without the model mucin degrader *Akkermansia muciniphila* DSM 2295932.

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. 8a,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. 8a). The growth of Roseburia 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. 8c).

To unveil the basis for the *Roseburia* growth, the proteomes of *R. hominis* and *R. inulinivorans*, both grown on glucose were compared with co-cultures of *Roseburia* and *A. muciniphila* grown on mucin. For *R. hominis* and *R. inulinivorans*, 31 and 93

---

**Fig. 3 Crystal structure of the GH136 lacto-N-biosidase from *E. ramulus* (ErLnb136).**

- **a-c** Overall structure and a semitransparent surface of *ErLnb136* consisting of an N-terminal domain (*ErLnb136$_I$, cyan-blue) and a C-terminal $\beta$-helix domain (*ErLnb136$_{II}$, green). The enzyme is shown in **a** a view orthogonal to the C-terminal $\beta$ helix domain, **b** the view of **a** rotated 180° and **c** a view along the axis of C-terminal $\beta$ helix domain, to highlight the interaction of *ErLnb136$_I$* and *ErLnb136$_{II}$*. **d** A molecular surface top view of the active site and a close up view **e** to illustrate the contribution of the *ErLnb136$_I$* 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 $\sigma$) of the LNB unit (yellow sticks) in the active site. The water (red sphere) mediated and direct hydrogen bonds that recognize the LNB are the yellow dashed lines. **d-f** The catalytic nucleophile (D575) and catalytic acid/base residue (D568) are labelled in red. **a-c** Disordered regions (residues 180–199 and 225–241) are shown as orange dotted lines.
proteins, including several CAZymes, were significantly upregulated (log₂ fold change > 2) relative to the glucose co-cultures. The transport protein RhLNBBP and RhGLnp112 from the R. hominis HMO locus (Fig. 1f) were the top 6th and 10th most upregulated proteins in the mucin proteome of R. inulinivorans, respectively, indicative of a role of this locus in cross-feeding on host glycans (Supplementary Fig. 8e). In R. inulinivorans, the corresponding enzymes RhLeᵃBP and RhGLnp112 were also significantly upregulated with log₂ fold changes of 2.77 and 4.71, respectively (Supplementary Fig. 8e). 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 (RiGH98), a putative α-galactosidase of GH36 and an α-1-fucosidase (GH29), which was the top fourth upregulated protein in the mucin proteome (Supplementary Fig. 8f). The upregulation of this locus suggested that R. inulinivorans possesses a functional machinery for directly accessing certain mucin oligomers. We expressed the fucosylated oligomers by the comparability of binding constants from these techniques. Both analyses were in independent duplicates (n = 2) and the Kᵦ values are reported with error bars representing the error of the fit to the binding isotherms. Growth experiments were performed as independent biological triplicates (n = 3) and triplicate HPAEC-PAD analyses were performed (one analysis/per biological replicate) whereby all HPAEC-PAD analyses yielded similar results.

proteins, including several CAZymes, were significantly upregulated (log₂ fold change > 2) relative to the glucose co-cultures. The transport protein RhLNBBP and RhGLnp112 from the R. hominis HMO locus (Fig. 1f) were the top 6th and 10th most upregulated proteins in the mucin proteome of R. inulinivorans, respectively, indicative of a role of this locus in cross-feeding on host glycans (Supplementary Fig. 8e). In R. inulinivorans, the corresponding enzymes RhLeᵃBP and RhGLnp112 were also significantly upregulated with log₂ fold changes of 2.77 and 4.71, respectively (Supplementary Fig. 8e). 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 (RiGH98), a putative α-galactosidase of GH36 and an α-1-fucosidase (GH29), which was the top fourth upregulated protein in the mucin proteome (Supplementary Fig. 8f). The upregulation of this locus suggested that R. inulinivorans possesses a functional machinery for directly accessing certain mucin oligomers. We expressed the fucosylated oligomers by the comparability of binding constants from these techniques. Both analyses were in independent duplicates (n = 2) and the Kᵦ values are reported with error bars representing the error of the fit to the binding isotherms. Growth experiments were performed as independent biological triplicates (n = 3) and triplicate HPAEC-PAD analyses were performed (one analysis/per biological replicate) whereby all HPAEC-PAD analyses yielded similar results.

The HMO utilization loci are prevalent in Roseburia. The HMO loci, defined by the co-occurrence of GH136 and GH112 genes, are conserved in five 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. As a reference signature for a central catabolic pathway, the presence of GH10 xylanase genes, compulsory for xylan utilization in R. intestinalis, was also analysed. Strikingly, the GH112 and GH136 HMO utilization
genes are about 2-3 fold more prevalent than the GH10 counterparts (Fig. 5a), indicative of the broader distribution of the HMO loci compared to the xylanase locus, which is mainly conserved in *R. intestinalis*. The GH136I and GH136II genes have a similar prevalence, which is ~30% lower than that of GH112. This overall trend is reiterated when we analyze individual species-level genome bins (SGBs), with some differences in the co-occurrence patterns of GH136 and GH112 genes (Fig. 5b). For example, while GH112 and GH136 have similar prevalence in *R. hominis* (SGB 4936), GH112 was 2.6 times more prevalent than GH136 in *R. inulinivorans* (SGB 4940). We analysed the organization of 818 loci, defined by the presence of a GH112 gene and at least one encoded subunit of the GH136, with a more stringent threshold (70% identity of the GH112 and GH136 sequences present in any of the 5 *Roseburia* reference genomes, see Supplementary Fig. 1). The gene clusters around the GH112 appeared to be SGBs-specific (Fig. 5c), indicative of diversity of the loci within the genus. Analysis of the most representative gene contexts for each SGB (Fig. 5d) shows that genes for ABC transporters, GH136, and transcriptional regulators were the most frequently co-occurring with GH112 genes, which offers a robust signature of the *Roseburia* HMO utilization loci (Fig. 5d) and validates their broad distribution. Additional CAZymes and carbohydrate metabolic genes were also frequently co-occurring in the vicinity of GH112 genes, suggesting that additional glycan utilization capabilities are clustered around the HMO loci.

**Fig. 5 The conservation and structure of HMO utilization loci in *Roseburia*.** a) Global abundance of GH112, GH136I, GH136II and GH10 xylanase genes in 4599 *Roseburia* genomes illustrating the broad occurrence and conservation of the HMO utilization apparatus. b) Heat map 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 genes within each SGB. This data shows the frequent co-occurrence of GH136 and GH112 genes, although some *Roseburia* strains encode only the GNB/LNB degrading GH112. c) Principal coordinate analysis of 818 *Roseburia* gene-landscapes defined stringently based on $\geq 70\%$ identity to the GH112 and GH136 with any of the five references *Roseburia* genomes displayed in Supplementary Fig. 1 and including 10 proteins up- and downstream of the GH112. d) The most frequently occurring gene landscapes in each *Roseburia* SGB, as anchored by aligning at the 3‘ terminal of GH112 genes. The gene landscape analyses provide a signature for the HMO utilization loci that are defined by 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\textsuperscript{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\textsuperscript{13,22}.

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 transitions of the assembly of the microbiota during infancy\textsuperscript{26,29} is available, but the underpinning mechanisms, especially during weaning, remain elusive. 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 may promote an early competitive adaptation advantage for Clostridiales that are associated with the healthy HGM and with the protection from metabolic and inflammatory disorders as well as colorectal cancer\textsuperscript{24-26,34}.

We uniquely demonstrate that key butyrate producing \textit{Roseburia} and \textit{Eubacterium} spp. grow on complex HMOs purified from mother’s milk and on defined HMO motifs (Fig. 1a–d). Proteomic analyses revealed two highly upregulated genetic loci that encode distant homologs to a lacto-N-biosidase from \textit{B. longum}\textsuperscript{28,35}, GNB/LNB phospholipases and ABC transporters in \textit{R. hominis} and \textit{R. inulinivorans}, (Fig. 1f–g and Supplementary Fig. 1). The \textit{R. hominis} locus (Figs. 1, 2 and 4, Supplementary Fig. 5e, Supplementary Tables 2, 4 and 5) supports growth on the HMO motifs LNT and LNB, whereas the \textit{R. inulinivorans} locus confers growth on more complex HMOs, e.g. single and double fucosylated versions of LNT (Figs. 2a and 4, Supplementary Figs. 2 and 5). The specialization on different, but partially overlapping, HMOs and related Lewis\textsuperscript{a/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, \textit{RhLn136} and \textit{ErLn136} are lacto-N-biosidases, whereas \textit{RILea/P136} displays an unprecedented specificity that requires a Fuc-\textalpha\textsubscript{-}1\textrightarrow4\textGlcNAc at the subsite −1 and accommodates additional fucosylation at the −2, and −2 subsites (Fig. 2, Supplementary Fig. 5a and Supplementary Table 4). The preference for fucosylation is consistent with an open active site effectuated by shortening of loops, (\textit{ErLn136}: Loop 1 AA 330-341, Loop 2 AA 520-543, Supplementary Fig. 6c), which allows the accommodation of bulky fucosylated substrates. Remarkably, the \textit{GH136} subunits (or domains in \textit{ErGH136-like enzymes}) are co-evolved with the \textit{GH136}\textsubscript{1} counterparts that possess the catalytic residues (Supplementary Fig. 6d).

Our stability (Supplementary Fig. 4c), structural (Fig. 3 and Supplementary Fig. 6), biochemical (Supplementary Fig. 4, Supplementary Table 4) and phylogenetic analyses (Supplementary Fig. 6d) affirm the crucial role of the \textit{GH136}\textsubscript{1} domain in the functionality of \textit{GH136} enzymes and provide compelling evidence to the association of the two \textit{GH136} domains. The sequence conservation of \textit{GH136}\textsubscript{1} and \textit{GH136}\textsubscript{1} was mapped on the structure of \textit{ErLn136}. Strikingly, highly conserved patches were identified across both domains (Supplementary Fig. 6e). Particularly, parts of the α4-α5 loop and of the α5 helix in \textit{ErLn136} that pack extensively onto \textit{ErLn136}\textsubscript{1} display globally conserved residues, together with the complementary co-conserved regions of \textit{ErLn136}\textsubscript{1} (Supplementary Fig. 6e).

Moreover, the surface of \textit{ErLn136}\textsubscript{1} is positively charged and apolar at the interface with \textit{ErLn136}\textsubscript{1}, which is notably different from the negative potential on the surface of the rest of the enzyme (Supplementary Fig. 6f) and complementary to the interface surface of \textit{ErLn136}\textsubscript{1}. These results highlight the co-evolution of \textit{GH136} subunits or domains.

ABC transporters are a determinant of uptake selectivity and competitive efficiency in both bifidobacteria\textsuperscript{17,31,36} and \textit{R. intestinalis}\textsuperscript{29}. The two SBPs of the ABC importers located in the HMO loci of \textit{R. hominis} and \textit{R. inulinivorans} were within the top 5 upregulated proteins in each proteome in response to HMO utilization (Fig. 1), underscoring the critical role of oligosaccharide transport in the competitive gut niche. The preferences of the SBPs and GHs encoded by these loci appear aligned to confer efficient uptake and subsequent catabolism of preferred substrates (Figs. 2 and 4, Supplementary Fig. 5, Supplementary Tables 2, 4 and 5). The LNB/GNB phosphorylases of \textit{GH112} are also conserved in the HMO loci (Supplementary Fig. 1). \textit{R. inulinivorans} possesses additional CAZymes, notably different fucosidases for degradation of internalized fucosylated-oligomers (Supplementary Fig. 1 and 5b–d). Based on the proteomic analyses and the biochemical data, we propose a model for the two distinct routes for uptake and depolymerisation of HMOs in \textit{Roseburia} and \textit{Eubacterium} (Fig. 6 and Supplementary Fig. 9).

Butyrate producing bacteria of the \textit{Roseburia-Eubacterium} group (Clostridiales order) are early colonizers of the infant gut\textsuperscript{5,37} and are prevalent members of the adult HGM\textsuperscript{22,23}.

The origin of this taxonomic group is enigmatic, but their presence in the human milk microbiome has been reported\textsuperscript{38,39}. Orthogonal transfer from mothers based on the identification of the same \textit{Roseburia} strains in mothers faeces, milk and the infant guts\textsuperscript{40} has also been proposed. \textit{R. intestinalis} type strains have been isolated from infant faeces\textsuperscript{41}, hinting the presence of this taxon before full transition to solid food.

We have previously shown that the abundance of distinct bifidobacteria in guts of breast-fed infants is strongly correlated to efficient ABC transporters that capture the 2′- and 3′-fucosyl-lactose HMOs with high affinity (\textit{K}\textsubscript{D} \approx 5\textmu M)\textsuperscript{17}. The strains possessing these genes, e.g. from \textit{Bifidobacterium longum} sub-species \textit{infantis}, are not detected after weaning, as opposed to counterparts adept at utilizing plant-derived glycans. By contrast, the same \textit{Clostridium} group XIVa strains that possess plant glycain utilization pathways\textsuperscript{29,42,43} retain HMO catabolic pathways. The simultaneous growth of \textit{R. hominis} on LNT and the cereal derived xylotetraose (Supplementary Fig. 2a–c) demonstrates this catabolic plasticity, which likely confers an additional competitive advantage during weaning, when the dominant fucosyl lactose specialized \textit{Bifidobacterium} community collapses due to sporadic supply of HMOs.

The loci that target HMOs also mediate cross-feeding on mucin or other glyco-conjugate oligomers, e.g. GNB from mucin and blood antigen structures, both captured efficiently by \textit{Roseburia} transport proteins (Fig. 4a, Supplementary Tables 2 and 3). This is consistent with the significant butyrate production measured in co-cultures of \textit{Roseburia} and \textit{A. muciniphila}\textsuperscript{32} (Supplementary Fig. 8c) and the upregulation of \textit{GH136}-containing loci in the mucin co-culture and HMO monocultures (Fig. 1 and Supplementary Fig. 8e). \textit{R. inulinivorans} possesses an extensive mucolytic machinery revealed by the upregulation of fucose and saccharic acid catabolism loci (Supplementary Fig. 3) as well as a blood group A and B- locus (Supplementary Fig. 8f–g, Supplementary Data 1) that allows the release of β-(1→4)-linked blood group oligomers found in mucin and glyco-lipids on the surfaces of enterocytes\textsuperscript{44,45}. This ability to access carbohydrates from mucin and host glyco-conjugates supports growth during periods of nutritional perturbations, which may increase the resilience of this taxonomic group.

Our bioinformatic analysis of the \textit{Roseburia} genomes establish that HMO utilization appears to be a core trait within \textit{Roseburia},
Fig. 6 Model for HMOs and related host glycan utilization by *Roseburia* and other Lachnospiraceae. In *R. hominis*, LNT, LNB and the mucin derived GNB are captured by RhlNBBP for uptake into the cytoplasm and LNT is subsequently hydrolysed to LNB. Both LNB and GNB are phosphorylated by RGlNbp112 into α-D-galactose-1-phosphate and the corresponding N-acetylated hexosamines GlcNAc and GalNAc, respectively. Lactose is likely hydrolysed by a canonical β-galactosidase. In *R. inulinivorans*, initial hydrolysis of HMOs or O-glycans from glyco-lipids/proteins occurs at the outer cell surface by RLεβ136, which has two C-terminal putative galactose-binding domains. The import of degradation products is mediated by the RLεβ136-associated ABC transporter. Fucosyl decorations are removed by the concerted activity of RIfuc95 and RIfuc29 before RGlNbp112 phosphorylates the resulting LNB or imported GNB into monosaccharides, as described in *R. hominis*. Galactose and galactose-1-phosphate products are converted via the Leloir pathway to glucose-6-phosphate and N-acetylated hexosamine sugars are converted to GlcNAc-6-phosphate before entering glycolysis. The pyruvate generated from glycolysis is partly converted to butyrate. *Roseburia* inhabits the outer mucus layer together with *A. muciniphila*. *R. inulinivorans* cross-feeds on sialic acid and accesses β-(1→4)-linked blood group A and B oligosaccharides from mucin and glyco-lipids/proteins via RGH98. Black solid arrows show enzymatic steps established or confirmed in this study. Black dotted arrows indicate steps based on literature. Grey dotted arrows indicate butyrate production by *R. hominis* and *R. inulinivorans* from mucin in co-culture with *A. muciniphila*. The glycan structure key is the same as in Fig. 1.

**Methods**

**Chemicals and carbohydrates.** 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 from Carbosynth and α-D-galactose and α-D-galactose-1-phosphate and the corresponding N-acetylated hexosamines GlcNAc and GalNAc, respectively. Lactose is likely hydrolysed by a canonical β-galactosidase. In *R. inulinivorans*, initial hydrolysis of HMOs or O-glycans from glyco-lipids/proteins occurs at the outer cell surface by Rlεβ136, which has two C-terminal putative galactose-binding domains. The import of degradation products is mediated by the Rlεβ136-associated ABC transporter. Fucosyl decorations are removed by the concerted activity of RIfuc95 and RIfuc29 before RGlNbp112 phosphorylates the resulting LNB or imported GNB into monosaccharides, as described in *R. hominis*. Galactose and galactose-1-phosphate products are converted via the Leloir pathway to glucose-6-phosphate and N-acetylated hexosamine sugars are converted to GlcNAc-6-phosphate before entering glycolysis. The pyruvate generated from glycolysis is partly converted to butyrate. *Roseburia* inhabits the outer mucus layer together with *A. muciniphila*. *R. inulinivorans* cross-feeds on sialic acid and accesses β-(1→4)-linked blood group A and B oligosaccharides from mucin and glyco-lipids/proteins via RGH98. Black solid arrows show enzymatic steps established or confirmed in this study. Black dotted arrows indicate steps based on literature. Grey dotted arrows indicate butyrate production by *R. hominis* and *R. inulinivorans* from mucin in co-culture with *A. muciniphila*. The glycan structure key is the same as in Fig. 1.

**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* (RGlNbp112). In detail, 100 mM Gal1P and 300 mM corresponding N-acetylated hexosamines (GlcNAc or GalNAc) in 50 mM MES, 150 mM NaCl, pH 6.5 were incubated with 10 μM RGlNbp112 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 (10,000×g, 30 min at 4 °C) to remove the enzyme. Supernatants were up concentrated by rotary evaporation and desaccharides were desalted in ultrapure water (milliQ) using a HiPrep Desalt column (GE Healthcare, Denmark) on an Akta avant chromatograph (GE Healthcare). Elution was monitored by measuring A235 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 guard column (4.6 × 250 mm) and a TSKgel Amide 80 guard column (4.6 × 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 analysed by thin layer chromatography (TLC) using 5 mM standards of GlcNAc, 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.** HMOs were purified from pooled human milk samples. Milk fat was separated by centrifugation (10,000 x 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 B-galactosidase from Rhamnothermus marinus (Sigma Aldrich) 1:1 (v/v) at 37 °C. The enzyme was precipitated with ethanol (as before) and the supernatant was con-
centrated 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% (v/v) formic acid (FA) and equilibrated with buffer A (with 4% (v/v) ACN, 0.05% (w/v) FA), which was also used to dilute the samples prior to loading. After sample loading, the columns were washed (6 column volumes of buffer A) to remove lactose and monosaccharides before elution. Oligosaccharides were eluted with 4% (v/v) ACN, 0.05% (w/v) FA. Eluted oligosaccharides were concentrated in a vacuum concentrator, freeze-dried and dissolved in milliQ prior to usage. Purity of HMOs was verified by high-performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD) on an ICS-5000 (Dionex) system with a 3 × 250 mm CarboPac PA200 column (ThermoScientific) and 10 µL injections. HMOs were eluted with a stepwise linear gradient of sodium acetate: 0-75 min of 0-150 mM and 25-35 min of 150-400 mM, at a flow rate of 0.35 mL min⁻¹ and a mobile phase of constant 0.1 mM NaOH. Standards (0.01-0.5 mM) of lactose, galactose and glucose in milliQ were used to quantify these residual sugars as described above. The analysis was performed in triplicates and the residual content of these sugars was <2% (w/w) of the purified HMO mixture.

Isolation and purification of porcine mucins. The commercial porcine gastric mucin (PGM) was further purified. In short, 20 g PGM was stirred for 20 h at 25 °C with 400 ml of an aqueous solution of 100 mM NaCl, 0.5% (w/v) bovine serum albumin (BSA) (from Sigma Aldrich) and 0.2% (w/v) sodium azide (Fluka) used as a preservative (1.2 after the first 2 h using 2 M NaOH). Insoluble residues were removed by centrifugation (10,000 g, 30 min at 4 °C) and soluble mucin was precipitated by the addition of 3 volumes of ice cold ethanol (95%) and incubation for 18 h at 4 °C. Precipitated mucin was dia lyzed 5 times against 200 volumes milliQ for 16 h at 4 °C, using a 50 kDa molecular weight cut off membrane (Spectra, VWR) and afterwards freeze dried. Porcine colonic mucin was isolated from five fresh pig colons from the slaughterhouse of Danish Crown (Horsens, Denmark). Pig colons were processed at site and immediately placed on dry ice to ensure quick cooling during transport. Colons were opened longitudinally and contents were removed mechanically and by washing with ice cold 0.9% (w/v) NaCl until no digesta was visible. Cleaned colons were opened longitudinally and content was removed mechanically and by freeze dried. Mucin was dialyzed 5 times against 200 volumes milliQ for 16 h at 4 °C, using a 50 kDa molecular weight cut off membrane (Spectra, VWR) and afterwards freeze dried.

Cloning, expression and purification of proteins. Open reading frames encoding proteins from R. hominis DSM 16839, R. inulinivorans DSM 16841 and E. ramulus DSM 16848 were cloned without signal peptide or transmembrane domain from genomic DNA using In-Fusion cloning (Takara) and the primers in Table S8 into mid-late exponential phase (OD₆₀₀ = 0.5–0.8) in four biological replicates. For R. hominis YCFA was supplemented with 0.5% (w/v) LNT or glucose and for R. inulinivorans 1% (w/v) HMOs or glucose was used as carbon source. Cells were harvested by centrifugation (5000 x g, 10 min at 4 °C) and supernatants for SCFA quantification at 24 and 48 h. Samples (2 mL) were collected for proteomics analyses after 16 h and for SCFA quantification after 24 and 48 h. Samples were immediately cooled on ice and cells were disrupted by centrifugation (10,000 x g, 10 min at 4 °C). For proteomics, cell pellets were washed twice with ice cold 0.9% (w/v) NaCl, resuspended in 60 µL lysis buffer and stored at −80 °C for proteomics analysis. Collection culture supernatants for SCFA quantification were sterile filtrated (0.45 µm filters) and stored at 80 °C for further analysis.

Sample preparation for mass spectrometry. Samples were processed using a previously established protocol. Cells were lysed by boiling (5 min 95 °C) followed by bead beating (3 mm beads, 30 Hz for 1 min) (TissueLyser II, Qiagen) and sonication bath (3 x 10 s at 4 °C) (Bioruptor, Diagenode). Lysates were centrifuged (14,000 x g, 10 min at 4 °C) and soluble protein concentrations were determined by a Bradford assay (Thermo Fisher Scientific). For digestion, 20 µg protein were diluted 1:0 with 50 mM HEPES, 10% (v/v) ACN, pH 8.5 and incubated with LysC (MS grade, Wako) in a ratio of 1:50 (LysC:protein) for 4 h at 37 °C. Subsequently, samples were diluted to 1:10 with 50 mM HEPES, 10% (v/v) ACN, pH 8.5 and further digested with trypsin (MS grade, Promega) in a ratio of 1:100 for 18 h at 37 °C. Next, samples were diluted 1:2 with 2% (w/v) trifluoroacetic acid (TFA) to quench enzymatic activity and peptides were processed for mass spectrometry using a protocol described stage 4 by Peters et al. [21]. Peptides from single strain cultures were desalted using three discs of C18 resin packed into a 200 µL tip and activated by successive loading of 40 µL of MeOH and 40% of 0.1% (v/v) TFA at 80 °C and equilibrated twice with 40% of 3% (v/v) ACN, 1% (v/v) FA before samples were loaded in steps of 50 µL. After loading, tips were washed three times with 100 µL 1% (w/v) TFA and peptides were eluted in two steps with 40 µL each of 40% (v/v) ACN, 0.1% (w/v) FA into a 0.5 mL Eppendorf LoBind tube. Peptides derived from co cultures were desalted and fractionated using strong cation exchange (SCX) chromatography filter plugs (3 M Empore). Per sample, 6 SCX discs were packed into a 100 µL groove and an aerobically grown culture (OD₆₀₀ ~0.01) of each condition was loaded (20 µL of ACN and then 80 µL of 0.2% (w/v) TFA). Samples were applied in 50 µL steps and tips were washed twice with 600 µL 0.2% (w/v) TFA. Subsequently peptides were stepwise eluted in 3 fractions with 60 µL of 125 mM NH₄OAc, 20% (v/v) ACN, 0.5% (w/v) FA, then with 60 µL of 225 mM NH₄OAc, 20% (v/v) ACN, 0.5% (w/v) FA and lastly with 5% (v/v) NH₄OH. 80% (v/v) ACN into 60 µL Eppendorf LoBind tubes. Eluted peptides were dried in an Eppendorf Speedvac (3 h at 60 °C) and reconstituted in 2% (v/v) ACN, 1% (w/v) TFA prior to mass spectrometry (MS) analysis.

LC-MS/MS. Peptides from biological triplicates of each culture condition were loaded on the mass spectrometer by reverse phase chromatography through an inline 50 cm C18 column (Thermo EasySpray ES803) connected to a 2 cm long C18 nanocolumn (Thermo Fisher Scientific). A nanoflow liquid chromatography (TFA) system. Peptides were eluted with a gradient of 48–48% (v/v) ACN, 0.1% (w/v) FA at 250 nL min⁻¹ over 260 min (samples from single strain cultures) or 140 min (SCX
fractionated samples from co-cultures) and analysed on a Q-Exactive instrument (Thermo Fisher Scientific) run in a data-dependent manner using a Top 10 method. MS data were collected at a resolution of 70,000 resolution, with an AGC target set to 3 × 10^6 ions or maximum injection time of 20 ms. Peptides were fragmented via higher-energy collision dissociation (normalized collision energy = 25). The intensity threshold was set to 1.7 × 10^6, dynamic exclusion to 60 s and ions with a charge state < 2 or known species were excluded. MS/MS spectra were acquired at a resolution of 17,500 and an AGC target of 1 × 10^6 ions or a maximum injection time of 60 ms. The scan range was limited from 300–1750 m/z.

**Protein label free quantification in bacterial co-cultures.** Proteome Discoverer versions 2.2 and 2.3 were used to process and analyse the raw MS data files and label free quantification was enabled in the processing and consensus steps. The spectra from single strains proteomics were matched against the proteome database of *R. hominis* DSM 1619 (ID: UP000008178) and *R. inulinivorans* DSM 16441 (ID: UP000003561) respectively, as obtained from Uniprot. The spectra from co-cultures experiments were searched against a constructed database consisting of the reference proteomes of the two *Roseburia* strains (as above) and *A. muciniphila* DSM 22959 (ID: UP000001031). For spectral searches, oxidation (M), deamidation (N, Q) and N-terminal acetylation were specified as dynamic modifications and cysteine carbamidomethylation was set as a static modification. Obtained results were filtered to a 1% FDR and protein quantitation was done by using the built-in Minora Feature Detector. For analysis of the label-free quantification data, proteins were considered present if at least two unique peptides (as defined in Proteome Discoverer) were identified and proteins to be identified in at least two out of the three samples from the culture condition per culture condition with high confidence.

Relative bacterial abundance in co-cultures was estimated based on strain unique peptides identified with Unipet version 4.0[^3]. To exclude peptides shared between closely related strains from the analyses, all peptide sequences quantified via Proteome Discoverer were imported into the Unipept web server and analysed with the settings Equal I and L and Advanced uniscan and cleavage handling options. The normalized sum of intensities of the resulting taxonomically distinctive peptides was then used for assessing relative abundances of each strain.

**Butyrate quantification.** Butyrate in culture supernatants was quantified by HPLC coupled to a refracting index detector (RID) and diode array detector (DAD) on an Agilent HP 1100 system (Agilent). Standards of butyric acid (0.09–50 mM) were prepared in 5 mM H2SO4 for peak identification and quantification. Samples from four biological replicates were analysed by injecting 20 µL of standard or prepared in 5 mM H2SO4 for peak identi

**Oligosaccharide uptake preference of *Roseburia spp.* **R. hominis** was grown anaerobically in 250 mL YCFA medium with 0.5% (w/v) of an equal mixture of xylose, galactose and LNT in biological triplicates. Samples (20 mL) were taken after 0, 3.5, 5.5, 6.5, 8, 9.5 and 24 h, diluted 10-fold in ice cold 100 mM NaOH and centrifuged (10 min at 5000 × g) before supernatants were stored at −20 °C until the HPAEC-PAD analysis. Standards of 0.5 mM xylose and LNT were prepared in 100 mM NaOH and used to identify corresponding peaks in the chromatograms. Aliquots (2 µL injections) were assayed 50 mM sodium phosphate buffer, 150 mM NaCl, 0.005% (v/v) Triton X-100 and OriginPro 2019b (OriginLab). Lacto-n-bioside specific activity of *RLaLb136* (1 µM) was measured as described above using 3.5 mM LNT. The specific activity was expressed in units (U) mg⁻¹ enzyme, where a unit is defined as the amount of enzyme that releases 1 µmol lactose min⁻¹ quantified as above.

**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 (3:5:2) (v/v) as mobile phase and sugars were visualized with 5-methylcoumarin (2:80:10) (v/v) and heat treatment except for *RLaLb136*. The TLC for the latter enzyme was performed in butanol:acetic acid:millilQ (2:1:1) (v/v) and developed with diphenylamine-phosphoric acid reagent[^27]. TLC analyses were performed from two independent biological duplicates (one analysis from each biological enzymatic reaction replicate).

**MALDI-TOF/MS.** MALDI-TOF/MS analysis of *RLaLb136* was according to[^56], following permethylation of oligosaccharides[^49]. For permethylation, hypoligosaccharides were reconstituted in 200 µL of anhydrous dimethylsulfoxide (DMSO) and mixed for 5 min with 250 µL of NaOH in DMSO and with 150 µL of iodomethane. Next, 2 mL of 5% (v/v) acetic acid was added followed by the addition of 2 mL of CH32. Subsequently, permethylated oligosaccharides were extracted in the organic phase, dried under a nitrogen stream at 40 °C before loading onto a pre-equilibrated Sep-pak C18 cartridges, washing with water and elution with 85% (v/v) acetonitrile. Eluted fractions were dried under nitrogen as before and stored at −20 °C until further use. After the enzymatic reaction permethylated products were dried, mixed with 2.5-dihydroxybenzoxacetic acid, and spotted onto the MALDI plate. For MALDI-TOF/MS analyses, a Bruker Autoflex III smartbeam in positive ion mode was used. Degradation products of *RLaLb136* and *ErLb136* were analysed 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). MALDI-TOF/MS analyses where performed from independent triplicates (one analysis from each biological enzymatic reaction replicate).
LC-MS² 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 was immobilized by dot blotting onto an immobilon-P PVDF membranes (Immobilon P membranes, 0.45 μm, Millipore, Billerica, MA). R/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 (Thermo Scientific) packed on top of a C18 Zip-tip (Millipore). Samples were eluted with 65% (v/v) ACN in 0.5% triethylamine in water (TFA, v/v), dried, resuspended in 10 μL of miliQ, frozen at −20 °C and stored until further analysis. The residual O-linked glycans (on the dot) were released by reducing the film with 100 μM DTT incubated at 37 °C. Samples were spotted onto a 0.5 μM 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 before60. The purified glycans were resuspended in 10 μL of miliQ and stored at −20 °C for further analysis. Released oligosaccharides from glycosphingolipids as a model substrate carrying blood group B (BS-2 and BS-624) were prepared as described above for a single incubation time of 2 h.

Purified samples were analysed 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 NaHCO₃, over 40 min at 10 μL min⁻¹. The eluted O-glycans were analysed on a LTQ mass spectrometer (Thermo Scientific) in negative-ion mode with an electrospray voltage of 3.5 kV, capillary voltage of −330 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 analysed. The data were processed using Xcalibur software (version 2.0, Thermo Scientific).

Oligosaccharide binding analysis. Binding of LNT, LNB, GNB, H type I trisaccharide, Le³ trisaccharide and Le³ tetrasaccharide to RlNh136-BBP was analyzed by surface plasmon resonance (SPR; Biacore T100, GE Healthcare). RlNh136-BBP, diluted in 10 mM NaOAc buffer pH 3.75 to 50 μg mL⁻¹, was immobilized on a CMS chip using a random amine coupling kit (GE Healthcare) to a final chip density of 3214 and 4559 response units (RU). Analysis comprised 90 s for association and 240 s for dissociation phase, respectively, at a flow rate of 30 μL min⁻¹. Sensograms were recorded at 25 °C in 2 mM sodium phosphate buffer, 150 mM NaCl, 0.05% (v/v) P20 (GE Healthcare), pH 6.5. Experiments were performed in duplicates (each consisting of a technical duplicate) in the range of integrated binding isotherms.

In further analysis, initial blast hits were filtered based on a 70% identity with any of the 5 conserved Roseburia reference genomes. Additionally, Roseburia genomes were considered only if they have a hit with GH112 gene and at least one subunit of the GH136 gene. The resulting 818 genomes were assigned into the respective Roseburia SGs, based on the assignment of Pasolli et al53. The retrieved genomes were used to analyse the gene landscape around the GH112 gene. The RAST server78 was used for gene annotation. Based on the annotation and coordinates of the genes, 11 genes upstream and downstream the GH112 were selected for genes landscape analysis. The most conserved gene neighborhood along each SGB was selected as the representative for each SGB. Principal component analysis was done based on the structure of the GH112-GH136 neighborhood, considering the present or absent of the genes on the gene landscape and as well its position on the loci. We used the function distmatrix with the “os” method from R to compute the distance matrix.

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

Reporting summary. Further information on research design is available in the Nature Research Supporting Information linked to this article.

Data availability. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD015045. The accession numbers for the atomic coordinates reported in this paper are PDB: 6KQS (Se-Met) (https://doi.org/10.2210/pdb6KQS/pdb) and 6KQT (native), see also Table S6. Mucin glycomics MS/MS data are summarized in Table S9 and raw data files are available upon request. Accession number of the cloned genes are provided in Figure legends. All other data are available from the corresponding author upon request. Source data are provided with this paper.

Received: 19 October 2019; Accepted: 11 June 2020; Published online: 03 July 2020
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. & Bäcked, F. Diet-microbiota interactions as modulators of human metabolism. Nature 535, 56–64 (2016).
3. Belkaid, Y. & Hand, T. W. Role of the microbiota in immunity and inflammation. Cell 157, 121–141 (2014).
4. Bokulich, N. A. et al. Antibiotics, birth mode, and diet shape microbiota 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. Bäcked, 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 gut microbiome and child growth—first 1000 days and beyond. Trends Microbiol. 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. Nat. Rev. Endocrinol. 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, 713–722 (2016).
13. Marchesi, J. R. et al. The gut microbiota and host health: a new clinical frontier. Gut 65, 330–339 (2016).
14. egt, K. 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, 323–335 (2012).
16. Gotth, A., Ojima, M. N. & Katayama, T. Minority species influences microbiota formation: the role of Bifidobacterium with extracellular glycosides in bifidus flora formation in breastfed infant guts. Microb. Biotechnol. 12, 259–264 (2019).
17. Sakakana, M. et al. Evolutionary adaptation in facosaclylsisupate uptake systems supports bifidobacteria-infant symbiosis. Sci. Adv. 5, eaaq7696 (2019).
18. Norin, E., Midvedt, T. & Björkstén, B. Development of faecal short-chain fatty acid pattern during the first year of life in Estonian and Swedish infants. Microb. Ecol. Health Dis. 16, 8–12 (2004).
19. Furusawa, Y. et al. Commensal microbe-derived butyrate induces the differentiation of colonial regulatory T cells. Nature 504, 446–450 (2013).
20. Bultman, S. J. Bacterial butyrate prevents atherosclerosis. Nat. Microbiol. 3, 1322–1333 (2018).
21. Corthesy, R. E. et al. Potential beneficial effects of butyrate in intestinal and extraintestinal diseases. World J. Gastroenterol. 17, 1519–1528 (2011).
22. Arumugam, M. et al. Enterotypes of the human gut microbiome. Nature 473, 174–180 (2011).
23. Louis, P. & Flint, H. J. Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. FEMS Microbiol. Lett. 24, 1–8 (2009).
24. Karlsson, F. H. et al. Symptomatic atherosclerosis is associated with an altered gut metagenome. Nat. Commun. 3, 1245 (2012).
25. Le Chatelier, E. et al. Richness of human gut microbiome correlates with metabolic markers. Nature 500, 541–546 (2013).
26. Wang, J. et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature 490, 55–60 (2012).
27. Vital, M., Karch, A. & Pieper, D. H. Colonic butyrate-producing communities in humans: an overview using omics data. mSystems. 2, e00130–17 (2017).
28. Yamada, C. et al. Molecular insight into evolution of symbiosis between breast-fed infants and a member of the human gut microbiome. FEMS Microbiol. Lett. 294, 1–8 (2009).
29. Skog, L. E. et al. Symptomatic atherosclerosis is associated with an altered gut metagenome. Nat. Commun. 3, 1245 (2012).
30. Le Chatelier, E. et al. Richness of human gut microbiome correlates with metabolic markers. Nature 500, 541–546 (2013).
31. Wang, J. et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature 490, 55–60 (2012).
32. Vital, M., Karch, A. & Pieper, D. H. Colonic butyrate-producing communities in humans: an overview using omics data. mSystems. 2, e00130–17 (2017).
33. Yamada, C. et al. Molecular insight into evolution of symbiosis between breast-fed infants and a member of the human gut microbiome. FEMS Microbiol. Lett. 294, 1–8 (2009).
34. Skog, L. E. et al. Symptomatic atherosclerosis is associated with an altered gut metagenome. Nat. Commun. 3, 1245 (2012).
35. Le Chatelier, E. et al. Richness of human gut microbiome correlates with metabolic markers. Nature 500, 541–546 (2013).
36. Wang, J. et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature 490, 55–60 (2012).
37. Vital, M., Karch, A. & Pieper, D. H. Colonic butyrate-producing communities in humans: an overview using omics data. mSystems. 2, e00130–17 (2017).
38. Yamada, C. et al. Molecular insight into evolution of symbiosis between breast-fed infants and a member of the human gut microbiome. FEMS Microbiol. Lett. 294, 1–8 (2009).
39. Skog, L. E. et al. Symptomatic atherosclerosis is associated with an altered gut metagenome. Nat. Commun. 3, 1245 (2012).
40. Le Chatelier, E. et al. Richness of human gut microbiome correlates with metabolic markers. Nature 500, 541–546 (2013).
41. Wang, J. et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature 490, 55–60 (2012).
42. Vital, M., Karch, A. & Pieper, D. H. Colonic butyrate-producing communities in humans: an overview using omics data. mSystems. 2, e00130–17 (2017).
43. Yamada, C. et al. Molecular insight into evolution of symbiosis between breast-fed infants and a member of the human gut microbiome. FEMS Microbiol. Lett. 294, 1–8 (2009).
44. Skog, L. E. et al. Symptomatic atherosclerosis is associated with an altered gut metagenome. Nat. Commun. 3, 1245 (2012).
Arakawa is thanked for managing the X-crystallography structural data and Drs.

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

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

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

Adams, P. D. et al. PHENIX: building new software for automated crystallographic structure determination. in Acta Crystallographica Section D: Biological Crystallography, Vol. 58, 1948–1954 (International Union of Crystallography, 2002).

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

Mishchukov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. Sect. D. Biol. Crystallogr. 53, 240–255 (1997).

Nielsen, H. Predicting secretory proteins with signalP. in Methods in Molecular Biology, Vol. 1611, 59–73 (Humana Press, New York, 2017).

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

Krogh, A., Larsson, B., Von Hejne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305, 567–580 (2001).

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

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

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

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

Huson, D. H. et al. Dendroscope: an interactive viewer for large phylogenetic trees. BMC Bioinformatics. 8, 460 (2007).

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

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

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

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

Acknowledgements

We would like to thank Dr. Ayaka Harada and Dr. Miki Senda, and the staff of the Photon Factory. We also wish to thank Tina Johansen for the technical help in performing the HPLC measurements for the quantification of butyrate. Dr. Takotshi Arakawa is thanked for managing the X-crystallography structural data and Drs. Fumihiko Sato and Kentro Ifuku are thanked for the technical support of MALDI-TOF/MS analysis. We also thank Prof. Tine Rask Licht for the use of the microplate reader for some of the growth experiments. We thank Dr. in spie Perrielle Neve for discussions on the taxonomic transitions of the early life microbiota. This study is funded by a PhD stipend for M.J.P. from the Technical University of Denmark, Kgs. Lyngby, Denmark. Additional funding was from a PhD scholarship from the Iraqi Ministry of Education to B.S. and Novo Nordisk Foundation–Copenhagen Bioscience PhD Program (grant NNF16CC0020896) supported C.A.S. Carlsberg Foundation is acknowledged for an ITC instrument grant (2011-01-0598) and DSC instrument grant (2013-01-0112). Additional funding is from the Swiss Light Source (grant numbers: 20181219 & 20181299) for the X-ray data collection and from the Swedish Research Council for funding the Swedish infrastructure for biological mass spectrometry (BioMS) used for the glycomics analysis.

Author contributions

Bacterial growth studies were performed by M.J.P. Proteomic analyses were done by M.J.P and E.S. Protein characterization was done by M.J.P and M.L.L. Enzymatic characterization of RLE imported was performed by A.G, To.K., M.S. and T.K. Mucin preparation was performed by B.S. and M.J.P. Mucin glycomics were performed by B.S., C.J. and N.G.K. Protein X-ray crystallography was performed by C.Y. and S.F. Metagenome analysis were performed by C.A-S and M.A. Experiments were designed by M.J.P and M.A.H. The paper was drafted by M.J.P and M.A.H and finalized with contributions of all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-17075-x.

Correspondence and requests for materials should be addressed to M.A.H.

Peer review information Nature Communications thanks Jolsef Vogtmann and the other, anonymous, reviewer for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020