ARTICLE

Check for updates

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

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

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.

[1] Department of Biotechnology and Biomedicine, Technical University of Denmark, 2800 Lyngby, Denmark. [2] Department of Biotechnology, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan. [3] Division of Biotechnology, Department of Chemistry, Lund University, 22100 Lund, Sweden. [4] Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen 2200, Denmark. [5] Graduate School of Biostudies, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan. [6] Faculty of Bioresources and Environmental Sciences, Ishikawa Prefectural University, Nonoichi, Ishikawa 921-8836, Japan. [7] Department of Medical Biochemistry, University of Gothenburg, 413 90 Gothenburg, Sweden. [8]These authors contributed equally: Michael Jakob Pichler, Chihaya Yamada ✉email: maha@bio.dtu.dk

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[4,5]. The HGM develops dynamically during infancy until a resilient adult-like community is formed after 2–3 years of life[6–8]. 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 life-long 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 *Clostridium* cluster IV and *Clostridium* cluster XIVa that includes the *Roseburia-Eubacterium* group (Lachnospiraceae family, Clostridiales order), which are abundant and prevalent in the adult HGM[22,23]. The abundance of *Roseburia* spp. is decreased in patients suffering from metabolic, inflammatory and cardiovascular diseases[24–27]. Although butyrate producers are established by the first year of life[27], the mechanisms underpinning their early appearance (and prevalence) remain unknown.

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.

Results

HMO loci in *Roseburia* and *Eubacterium*. Our aim was to investigate the HMO utilizsation 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) in the Carbohydrate Active enZyme (CAZyme) 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). 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 CAZyme 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 ($p < 4.0 \times 10^{-4}$) 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 ($\mu_{max} = 0.30 \pm 0.01$ 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 ($\mu_{max} = 0.22 \pm 0.02$ h^{-1}), its LNB unit ($\mu_{max} = 0.16 \pm 0.01$ h^{-1}) and the mucin-derived galacto-*N*-biose (GNB) ($\mu_{max} = 0.21 \pm 0.02$ h^{-1}) (Fig. 1a). Growth on LNT was also shared by the taxonomically related *Eubacterium ramulus* 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 favourite 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$_2$ fold change > 2). These differential proteomes were dominated by carbohydrate metabolism proteins, especially products of two loci
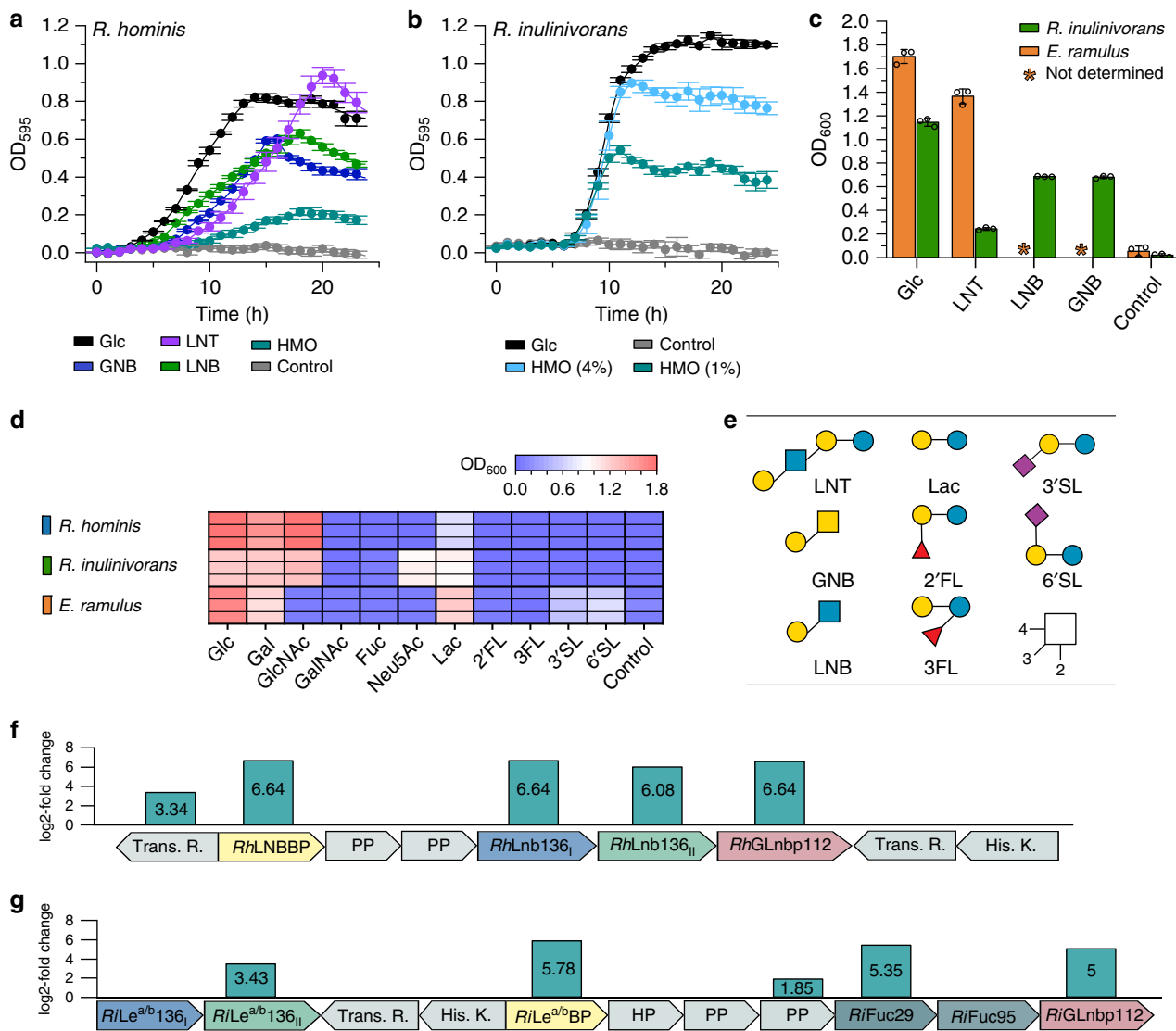
Fig. 1 Growth of Roseburia and Eubacterium spp. on HMOs and upregulation of HMOs utilization loci in Roseburia. Growth curves of R. hominis (**a**) and R. inulinivorans (**b**) on glucose, LNT, GNB, LNB, and/or purified HMOs from mother's milk compared to no-carbon source controls over 24 h. **c** Growth 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, 3FL, 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 (**f**) and RiLe$^{a/b}$BP (**g**)); ABC transporter permease protein (PP); hypothetical proteins (HP); Glycoside hydrolase 136 (RhLnb136$_I$, RhLnb136$_{II}$ (**f**) and RiLe$^{a/b}$136$_I$, RiLe$^{a/b}$136$_{II}$ (**g**)); Glycoside hydrolase 112 (RhGLnbp112 (**f**) and RiGLnbp112 (**g**)); Glycoside hydrolase 29 (RiFuc29 (**g**)); Glycoside hydrolase 95 (RiFuc95 (**g**)) and histidine kinase sensory protein (His. K.) The proteomic analyses (**f–g**) were 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). Source data are provided as a Source data file labelled with the corresponding figure number and panel definition.

(henceforth referred to as HMO utilization 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 (RhLNBBP) and R. inulinivorans (RiLe$^{a/b}$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).

Diverse GH136 enzymes mediate initial HMO degradation. The homologs RhLnb136$_I$ (LnbY in B. longum) and RhLnb136$_{II}$ (LnbX that harbours the catalytic residues in B. longum) were highly co-upregulated in the LNT proteome of R. hominis (Fig. 1f). Both, RhLnb136$_I$ and RhLnb136$_{II}$ lack a predicted
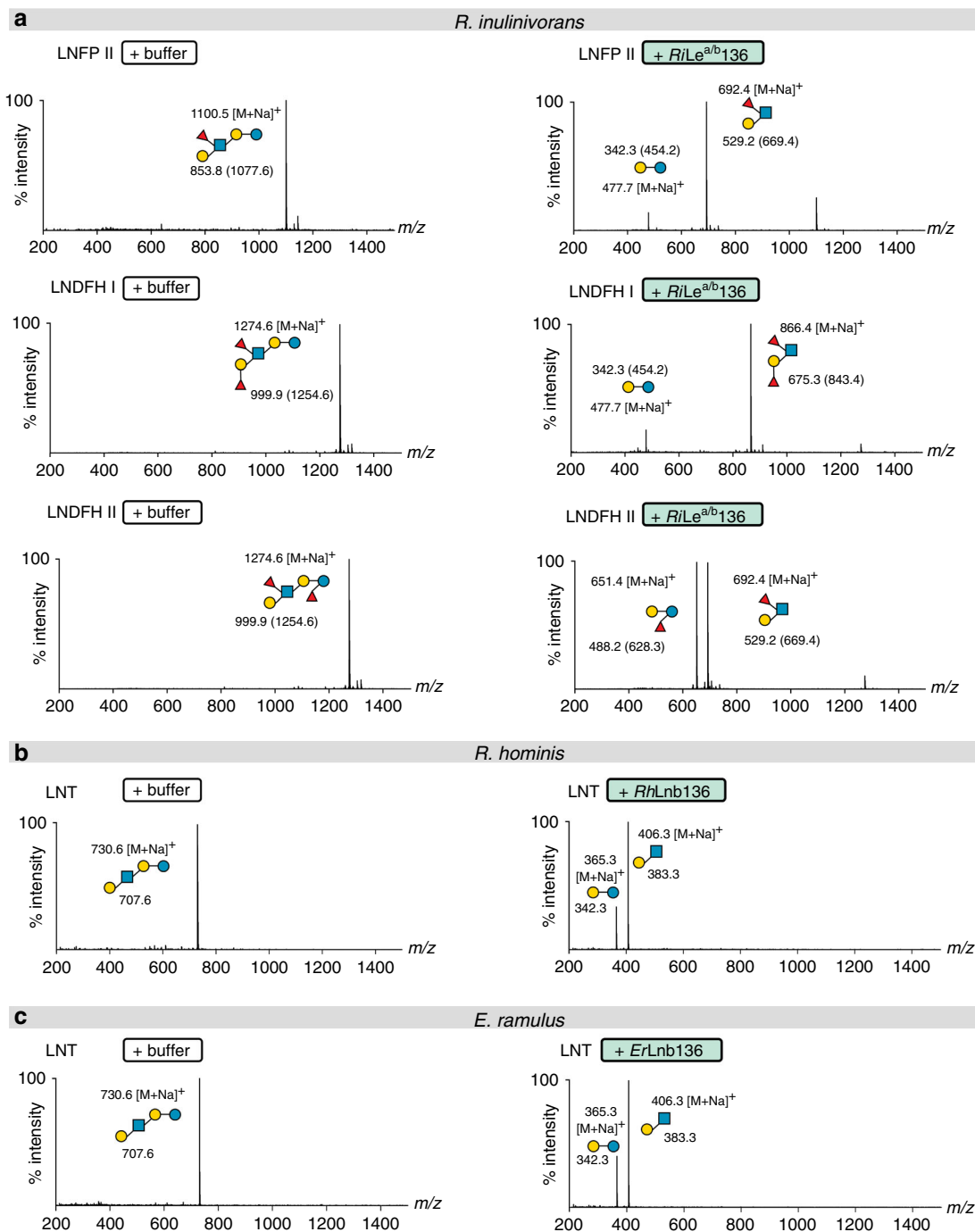
Fig. 2 Specificities of GH136 enzymes that mediate the HMO degradation. a Activity of *Ri*Le^a/b^136 on fucosylated HMOs. **b** Activity of *Rh*Lnb136 on LNT. **c** Activity of *Er*Lnb136 on LNT. **a–c** The hydrolysates were analysed by MALDI-ToF MS without (**b**, **c**) or with **a** permethylation. **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.

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*Lnb136$_I$ and *Rh*Lnb136$_{II}$ 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*Lnb136$_I$ and *Rh*Lnb136$_{II}$ 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 catabolism role of the locus.

*Ri*GH136, the GH136 homolog from the HMO-upregulated locus in *R. inulinivorans* was predicted to be extracellular, with a signal peptide in *Ri*GH136$_{II}$ that also possesses two C-terminal putative carbohydrate binding modules (Supplementary Fig. 8a) and a predicted N-terminal transmembrane domain in *Ri*GH136$_I$. Co-expression of *Ri*GH136$_I$ and *Ri*GH136$_{II}$, lacking the transmembrane domain and signal peptide respectively, resulted in an active enzyme with an unprecedented specificity. This enzyme (*Ri*Le$^{a/b}$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 *Ri*Fuc29 and *Ri*Fuc95 that act on α-(1 → 4) and α-(1 → 2)-linked ʟ-fucosyl, respectively mediates the complete defucosylation of putative products of *Ri*GH136, Le^b tetraose, Le^a triose and H triose type I (Supplementary Fig. 5b–d). Initial defucosylation by *Ri*Fuc29 is required for releasing the 1 → 2 linked ʟ-fucosyl in Le^b tetraose by *Ri*Fuc95. Finally, we showed that the GH112 from *R. inulinivorans* (*Ri*GLnbp112) 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 *Rh*Lnb136 and *Ri*Le$^{a/b}$136, without success. Hence, we turned our attention to the taxonomically related *E. ramulus*, which has a GH136 locus similar to the one in *R. hominis*, except for a substitution of the GH112 phosphorylase with a GH42 β-galactosidase gene (Supplementary Fig. S1) that may confer hydrolysis of the LNB/GNB units[30]. *E. ramulus* and *R. hominis* also shared similar growth profiles on LNT (Fig. 1), which was consistent with the presence of a functional GH136. The N-terminal (*Er*Lnb136$_I$) and the C-terminal (*Er*Lnb136II) regions of the *E. ramulus* GH136 share homology to *Rh*Lnb136$_I$ and *Rh*Lnb136$_{II}$ (Supplementary Fig. 1 and 4), respectively, suggestive of the fusion of *Er*Lnb136$_I$ and *Er*Lnb136$_{II}$ to form an active enzyme (*Er*Lnb136). This is supported by the identical specificity and similar catalytic efficiencies of *Er*Lnb136 and *Rh*Lnb136 (Supplementary Table 4). Moreover, intimate interaction of the two *Er*Lnb136 domains is consistent with the cooperative unfolding profile of the enzyme (Supplementary Fig. 4b). These data justified the use *Er*Lnb136 to study the architecture of the two subunits/domains compulsory for activity within GH136. The crystal structures of selenomethionine (SeMet)-labelled and native *Er*Lnb136 were determined at 1.4 and 2.0 Å resolution, respectively (Supplementary Table 6). The C-terminal catalytic domain (*Er*Lnb136$_{II}$, from AA 242-663) assumes a β helix fold (Fig. 3) similar to the bifidobacterial homolog LnbX (Supplementary Table 7). The LNB molecule bound in the active site is recognized by ten potential hydrogen bonds and aromatic stacking of the Gal unit onto W548 (Fig. 3f and Supplementary Fig. 6a). Interestingly, the GlcNAc sugar ring of LNB in *Er*Lnb136 adopts an ^4E conformation ($φ = 232°$ and $ψ = 68°$) with the O1-OH in a pseudoaxial position to form a direct hydrogen bond with the acid/base catalyst (D568) (Supplementary Fig. 6a). Moreover, the D575 O^{δ2} of the nucleophile is positioned appropriately for nucleophilic attack on the anomeric carbon of the GlcNAc at 3.2 Å (Fig. 3f).

The N-terminal domain (*Er*Lnb136$_I$, from AA 7-224) consists of 8 α-helices (α1-α8) (Fig. 3a–c) and assumes a previously unknown fold, stabilized by the central helix α1. The structurally most related protein to *Er*Lnb136$_I$, a peptidyl-prolyl cis-trans isomerase with a chaperone activity from *Helicobacter pylori* (5EZ1), shares weak structural similarity restricted to helices α6 and α7 (Supplementary Fig. 6b, Supplementary Table 7). The *Er*Lnb136$_I$ domain embraces the sides and back of the β helix domain (Fig. 3a–c). These extensive inter-domain interactions (solvent inaccessible interface ≈1618 Å²), stabilize the protein structure with $ΔG = -17$ kcal mol^{-1}. Remarkably, the α6-α7 loop of *Er*Lnb136$_I$ forms a part of the active site with the solvent accessible sidechain of Y145 positioned near the active site (5.7 Å to the GlcNAc O1 atom of LNB) (Fig. 3d, e). The Y145A mutant showed a 4.9-fold higher K_M (Supplementary Table 4, Supplementary Fig. 4c), suggesting that this residue contributes to substrate interactions, possibly at the +1 subsite.

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 (LNB-binding protein, *Rh*LNBBP) shows preference to LNB followed by GNB and LNT, suggestive of the uptake and intracellular degradation of these ligands by *Rh*GLnbp112 and *Rh*Lnb136 as described above. By contrast, fucosyl-decorated Lewis b (Le^b) tetraose and Lewis a (Le^a) triose were the preferred ligand of the Le$^{a/b}$ binding protein (*Ri*Le$^{a/b}$BP) 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 *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. The specificity of *Ri*Le$^{a/b}$BP is highlighted by the lack of affinity for lacto-*N*-neotetraose (LN*n*T), blood group A antigen triose (A triose), lactose and 2′-fucosyllactose (2′-FL). These findings show that the products of *Ri*Le$^{a/b}$136 are the preferred ligands for *Ri*Le$^{a/b}$BP, consistent with the extracellular degradation of fucosylated pentaose and hexaose HMOs and uptake of their products by the ABC transporter. An equimolar mixture of Le^b, Le^a and H-triose type I oligomers promoted the growth of *R. inulinivorans* to a similar final OD_{600} as glucose (Fig. 1b, c and Fig. 4a). The uptake profiles of these ligands reflected the preference of *Ri*Le$^{a/b}$BP, consistent with uptake by the associated transporter (Fig. 4). This was also in accord with the utilization of larger fucosylated HMO structures observed during growth on purified HMOs from mother's milk (Supplementary Fig. 2a–c). Notably, no uptake of LNT was observed, which is in excellent agreement with the poor growth (Fig. 1c) and with the lack of detectable binding to LNT by *Ri*Le$^{a/b}$BP (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 *Ri*Le$^{a/b}$BP 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 HGM members. To evaluate possible mechanisms of cross-feeding we compared *Roseburia*
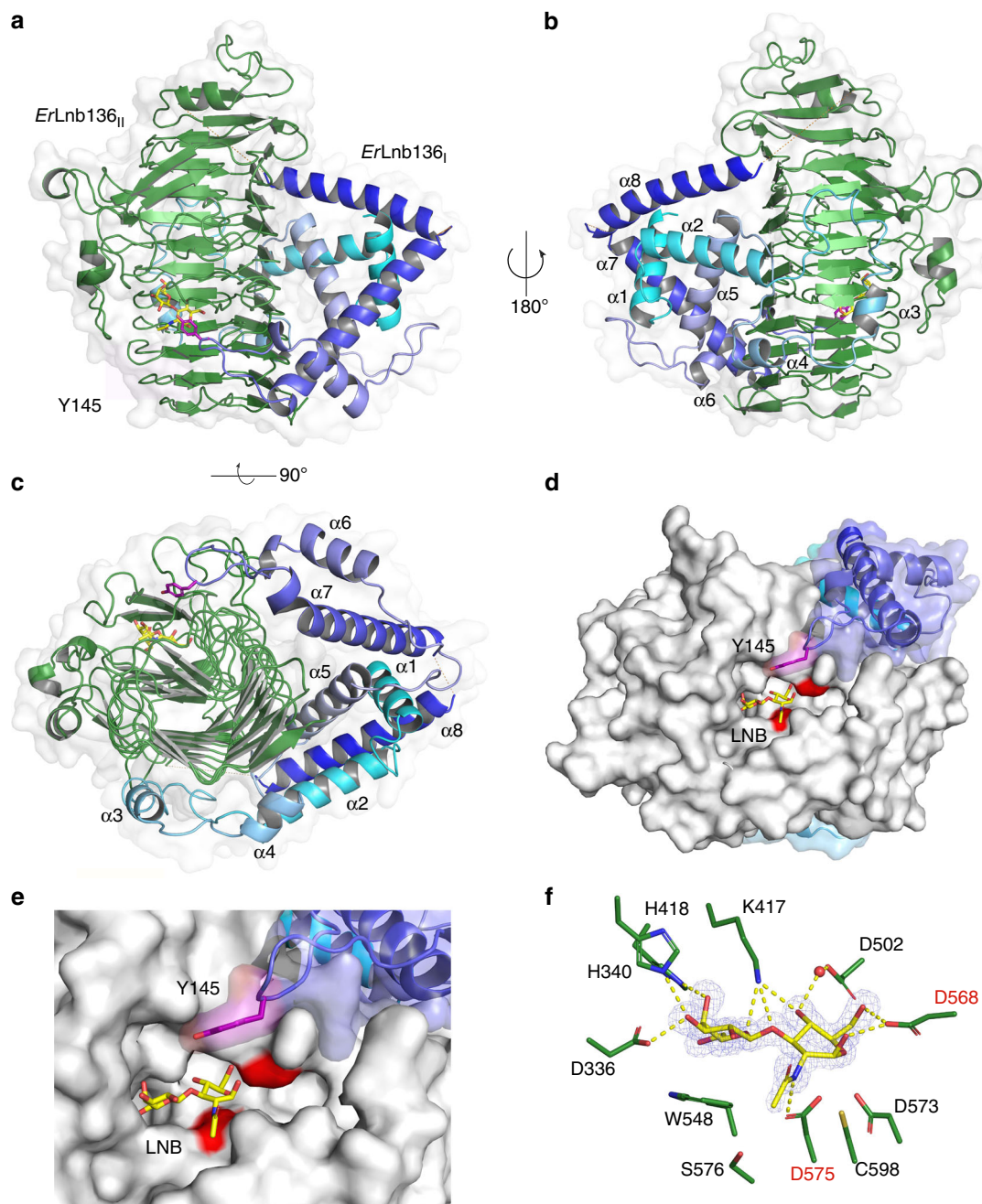
Fig. 3 Crystal structure of the GH136 lacto-*N*-biosidase from *E. ramulus* (*Er*Lnb136). **a–c** Overall structure and a semitransparent surface of *Er*Lnb136 consisting of an N-terminal domain (*Er*Lnb136$_I$, cyan-blue) and a C-terminal β-helix domain (*Er*Lnb136$_{II}$, green). The enzyme is shown 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 *Er*Lnb136$_I$ and *Er*Lnb136$_{II}$. **d** A molecular surface top view of the active site and a close up view **e** to illustrate the contribution of the *Er*Lnb136$_I$ 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 σ) 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.

growth on mucin with and without the model mucin degrader *Akkermansia muciniphila* DSM 22959[32].

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
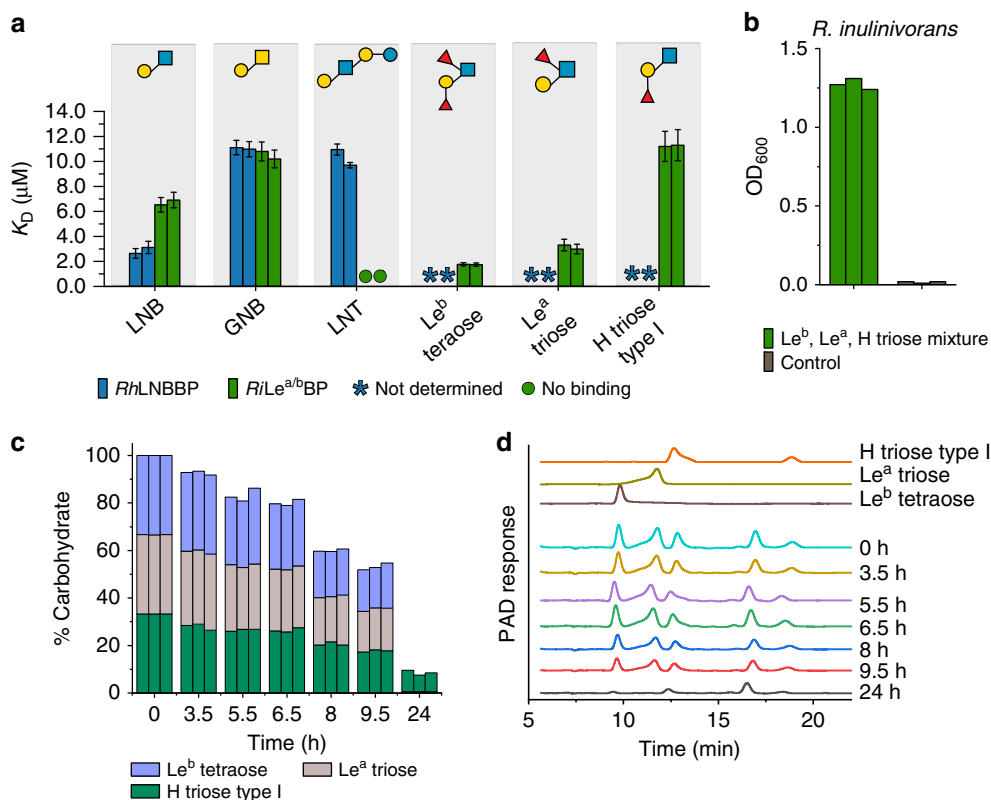
Fig. 4 *Roseburia* transport proteins mediate capture of HMOs and related host derived oligosaccharides. a Binding analysis of HMOs and host derived oligosaccharides to *Rh*LNBBP and *Ri*Le^a/b^BP. **b**, **c** Growth and uptake preference of *R. inulinivorans* on an equimolar mixture of Le^b tetraose, Le^a triose and H triose type I. **b** Growth level of *R. inulinivorans* on an equimolar mixture of Le^b tetraose, Le^a triose and H triose type I within 24 h including a no-carbon control. **c** Time course of the relative percentages of Le^b tetraose, Le^a triose and H triose type I in culture supernatants from **b** based on HPAEC-PAD analyses presented in **d**. **d** Representative HPAEC-PAD chromatograms showing time course analysis of culture supernatants of *R. inulinivorans* grown on YCFA media supplemented with 1.5 mM Leb tetraose, 1.5 mM Le^a triose and 1.5 mM H triose type I. Binding affinities **a** of *Rh*LNBBP were determined by isothermal titration calorimetry (ITC) while binding affinities of *Ri*Le^a/b^BP were determined by surface plasmon resonance (SPR) due to low availability of the ligands and justified by the comparability of binding constants from these techniques[17,31]. Both analyses were in independent duplicates ($n = 2$) and the K_D values are reported with error bars representing the error of the fit to the binding isotherms. Growth experiments **b** were performed as independent biological triplicates ($n = 3$) and triplicate HPAEC-PAD analyses **c**, **d** were performed (one analysis/per biological replicate) whereby all HPAEC-PAD analyses yielded similar results.

proteins, including several CAZymes, were significantly upregulated (log$_2$ fold change > 2) 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, indicative of a role of this locus in cross-feeding on host glycans (Supplementary Fig. 8e). In *R. inulinivorans*, the corresponding enzymes *Ri*Le^a/b^BP and *Ri*GLnbp112 were also significantly upregulated with log$_2$ 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 (*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 (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 predicted extracellular *Ri*GH98 and demonstrated release of blood group A and B oligomers from mucin and related *O*-glycans (Supplementary Fig. 8g–h, Supplementary Data 1). 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). The poor growth of *R. inulinivorans* on mucin in the absence of *A. muciniphila* suggests that the latter bacterium enables cross-feeding by the release of sialic acid, as *R. inulinivorans* lacks sialidases, which together with the ability of *Ri*GH98 to access blood group A and B oligomers in mucin substrates, may support a better co-growth in the co-culture. These findings are consistent with the role of HMO utilization machinery and additional functional operons in supporting co-growth with *A. muciniphila* on mucin.

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[33]. As a reference signature for a central catabolic pathway, the presence of GH10 xylanase genes, compulsory for xylan utilization in *R. intestinalis*[29], was also analysed. Strikingly, the GH112 and GH136 HMO utilization
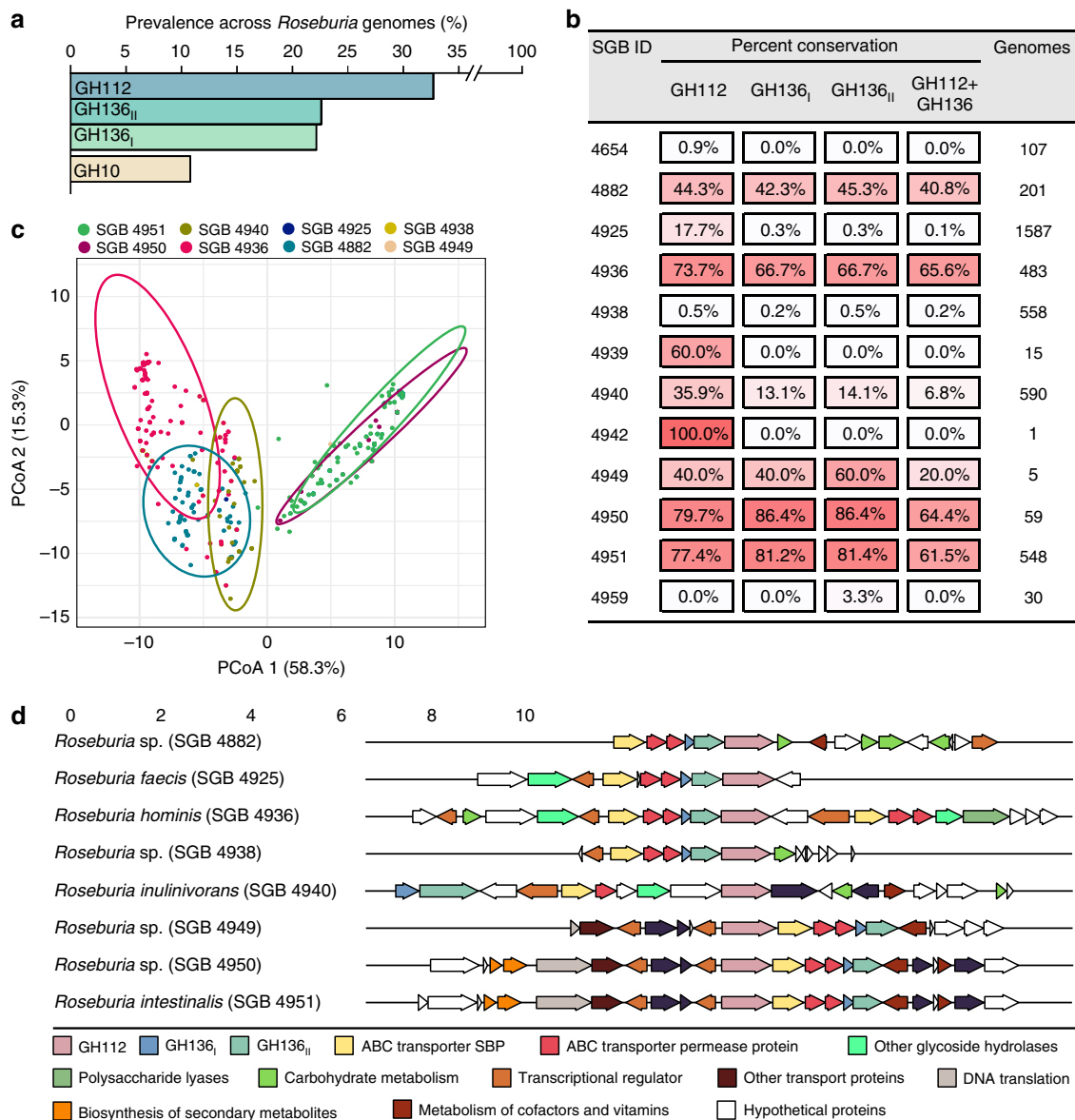
Fig. 5 The conservation and structure of HMO utilization loci in Roseburia. **a** Global abundance of GH112, GH136$_I$, GH136$_{II}$ and GH10 xylanase genes in 4599 *Roseburia* genomes illustrating the broad occurrence and conservation of the HMO utilization apparatus. **b** Heat map showing the segregation 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 ≥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.

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 GH136$_I$ and GH136$_{II}$ 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.

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 transitions of the assembly of the microbiota during infancy[6–8] 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[24–26,34].

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–d). Proteomic analyses revealed two highly upregulated genetic loci that encode distant homologs to a lacto-N-biosidase from B. longum[28,35], GNB/LNB phosphorylases and ABC transporters in R. hominis and R. inulinivorans, (Fig. 1f–g and Supplementary Fig. 1). The 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 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[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, RhLnb136 and ErLnb136 are lacto-N-biosidases, whereas RiLe[a/b]136 displays an unprecedented specificity that requires a Fuc-α-$(1 \rightarrow 4)$-GlcNAc 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 to fucosylation is consistent with an open active site effectuated by shortening of loops, (ErLnb136: Loop 1 AA 330-341, Loop 2 AA 520-543, Supplementary Fig. 6c), which allows the accommodation of bulky fucosylated substrates. Remarkably, the GH136$_I$ subunits (or domains in ErGH136-like enzymes) are co-evolved with the GH136$_{II}$ 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 GH136$_I$ domain in the functionality of GH136 enzymes and provide compelling evidence to the association of the two GH136 domains. The sequence conservation of GH136$_I$ and GH136$_{II}$ was mapped on the structure of ErLnb136. 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 ErLnb136$_I$ that pack extensively onto ErLnb136$_{II}$ display globally conserved residues, together with the complementary co-conserved regions of ErLnb136$_{II}$ (Supplementary Fig. 6e). Moreover, the surface of ErLnb136$_I$ is positively charged and apolar at the interface with ErLnb136$_{II}$, 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 ErLnb136$_{II}$. These results highlight the co-evolution of GH136 subunits or domains.

ABC transporters are a determinant of uptake selectivity and competitiveness in both bifidobacteria[17,31,36] and R. intestinalis[29]. The two SBPs of the ABC importers located in the HMO loci of R. hominis and 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, 3, 4 and 5). The LNB/GNB phosphorylases of GH112 are also conserved in the HMO loci (Supplementary Fig. 1). 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 Roseburia and Eubacterium (Fig. 6 and Supplementary Fig. 9).

Butyrate producing bacteria of the Roseburia-Eubacterium group (Clostridiales order) are early colonizers of the infant gut[6,8,37] and are prevalent members of the adult HGM[22,23].

The origin of this taxonomic group is enigmatic, but their presence in the human milk microbiome has been reported[38,39]. Orthogonal transfer from mothers based on the identification of the same Roseburia strains in mothers faeces, milk and the infant guts[40] has also been proposed. R. intestinalis type strains have been isolated from infant faeces[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 ($K_D \approx 5\,\mu M$)[17]. The strains possessing these genes, e.g. from Bifidobacterium longum subspecies infantis, are not detected after weaning, as opposed to counterparts adept at utilizing plant-derived glycans. By contrast, the same Clostridium group XIVa strains that possess plant glycan utilization pathways[29,42,43] retain HMO catabolic pathways. The simultaneous growth of 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 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 Roseburia transport proteins (Fig. 4a, Supplementary Tables 2 and 3). This is consistent with the significant butyrate production measured in co-cultures of Roseburia and A. muciniphila[32] (Supplementary Fig. 8c) and the upregulation of GH136-containing loci in the mucin co-culture and HMO monocultures (Fig. 1 and Supplementary Fig. 8e). R. inulinivorans possesses an extensive mucolytic machinery revealed by the upregulation of fucose and sialic 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-$(1 \rightarrow 4)$-linked blood group oligomers found in mucin and glyco-lipids on the surfaces of enterocytes[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 Roseburia genomes establish that HMO utilization appears to be a core trait within Roseburia,
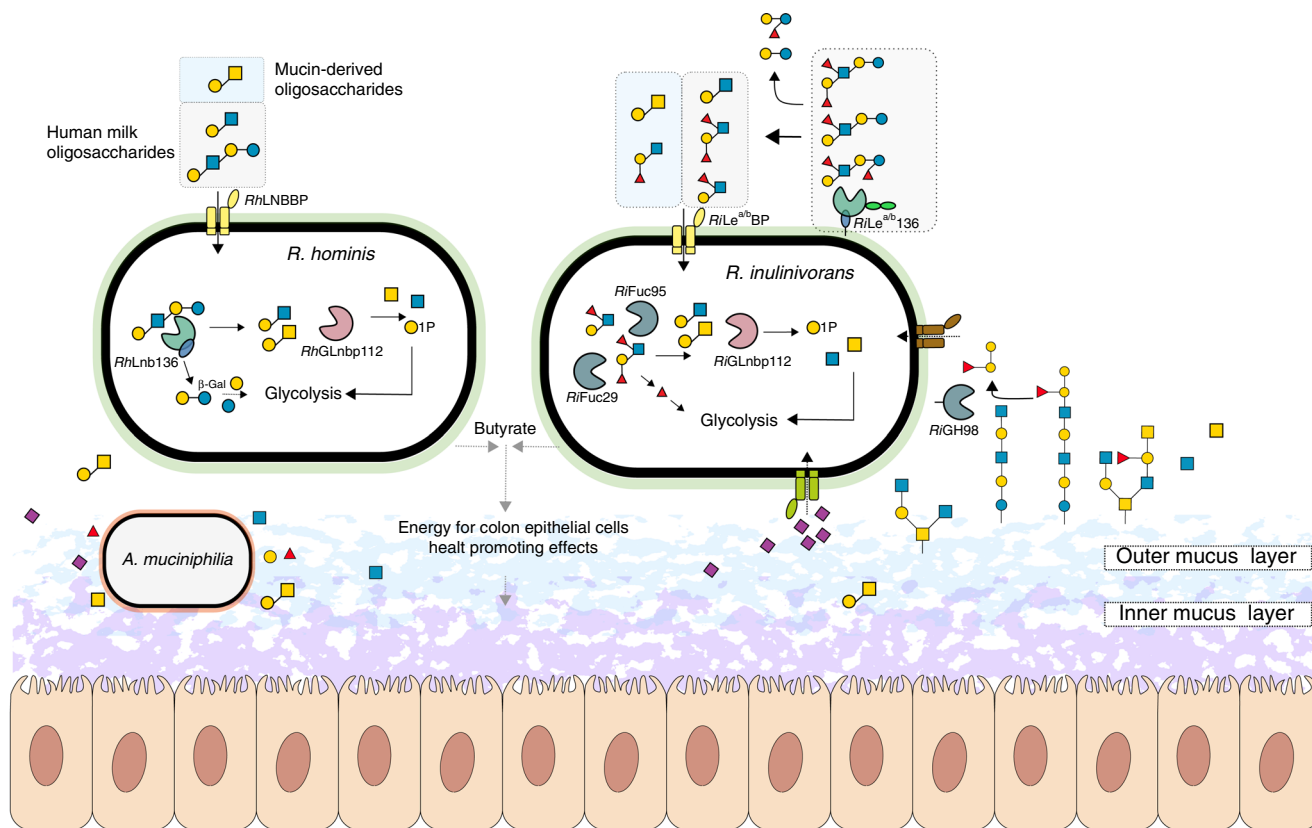
Fig. 6 Model for HMOs and related host glycan utilization by *Roseburia* and other Lachnospiraceae. In *R. hominis*, LNT, LNB and the mucin derived GNB are captured by *Rh*LNBBP for uptake into the cytoplasm and LNT is subsequently hydrolysed to LNB. Both LNB and GNB are phosphorolyzed by *Rh*GLnbp112 into α-D-galactose-1-phosphate and the corresponding *N*-acetylhexosamines 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 *Ri*Le$^{a/b}$136, which has two C-terminal putative galactose-binding domains. The import of degradation products is mediated by the *Ri*Le$^{a/b}$BP-associated ABC transporter. Fucosyl decorations are removed by the concerted activity of *Ri*Fuc95 and *Ri*Fuc29 before *Rh*GLnbp112 phosphorolyzes 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*-acetylhexosamine sugars are converted to GlcNAc-6-phosphate before entering glycolysis. The pyruvate generated from glycolysis is partly converted to butyrate[46]. *Roseburia* inhabits the outer mucus layer[47] together with *A. muciniphilia*. *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 *Ri*GH98. 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. muciniphilia*. The glycan structure key is the same as in Fig. 1.

based on the ubiquitous presence of loci harbouring GH112 and GH136 genes (Fig. 5). The occurrence of SGBs that exclusively possess GH112 genes (e.g. SGB 4939, Fig. 5b) suggests that distinct strains are secondary degraders that cross-feed on released simple substrates, e.g. LNB and GNB. By contrast, the co-occurrence of GH112 and GH136 genes (Fig. 5b) offers a signature for primary degraders that are able to access more complex glycans from HMOs or host glyco-conjugates.

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

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 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 pooled human milk samples were purchased from Hvidøvre hospital

(Hvidøvre, Denmark). All chemicals were of analytical grade unless otherwise stated.

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 (10,000×*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 × 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 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. HMOs were purified from pooled human milk samples[48,49]. Milk fat was separated by centrifugation (10,000 × *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 *Kluyveromyces 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% (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 oligosaccharides were eluted with 40% (v/v) ACN, 0.05% (w/v) FA. Eluted oligosaccharides were concentrated in a speed 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 (Theromofisher), a 3 × 50 mm CarboPac guard column (Theromofisher) and 10 µL injections. HMOs were eluted with a stepwise linear gradient of 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 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[50]. In short, 20 g PGM was stirred for 20 h at 25 °C in 20 mM phosphate buffer, 100 mM NaCl, pH 7.8 (adjusted to pH 7.2 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 (99%) and incubation for 18 h at 4 °C. Precipitated 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.

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 content was removed mechanically and by washing with ice cold 0.9% (w/v) NaCl until no digesta was visible. Cleaned luminal surface was quickly dried with absorptive paper and the mucosa was scraped off with a blunt metal spatula and subsequently transferred into a pre-cooled glass beaker whereby visible fat was removed and discarded. Mucin was then purified as previously described[51]. Isolated mucin was immersed in 10 volumes extraction buffer (10 mM sodium phosphate buffer, 6 M guanidine hydrochloride (GuHCl), 5 mM ethylenediaminetetraacetic acid (EDTA), 5 mM N-ethylmaleimide, pH 6.5) and gently stirred overnight at 4 °C. Soluble impurities and floating fat were separated by centrifugation (10,000×g, 30 min at 4 °C), pelleted mucin was dissolved in 10 volumes extraction buffer and incubated for 3 h at room temperature again. Soluble impurities were removed by centrifugation as described before. Short incubation (3 h) extraction steps were repeated 7 times until the supernatant was clear for at least two repeated extractions. Afterwards insoluble mucin was solubilized by reduction in 0.1 M Tris, 6 M GuHCl, 5 mM EDTA, 25 mM dithiotreitol (DTT) pH 8, for 5 h at 37 °C and subsequent alkylation through the addition of 65 mM iodoacetamide and incubation in the dark for 18 h at 4 °C. Soluble mucin was dialyzed 6 times against 200 volumes milliQ using a 50 kDa MWCO dialysis bag for 6 h at 4 °C and 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 15684 were cloned without signal peptide or transmembrane domain from genomic DNA using In-Fusion cloning (Takara) and the primers in Table S8 into the EcoRI and NcoI restriction sites of the corresponding plasmids, to encode proteins with either a cleavable N- or C- terminal His$_6$ tag. The pETM 11 plasmid was used (from G. Stier, EMBL, Center for Biochemistry, Heidelberg, Germany)[52], except for *RHOM_04110* (*Rh*Lnb136$_I$) and *ROSEINA2194_01899* (*Ri*Le$^{a/b}$136$_I$) which were cloned into pET15b (Novagen). Recombinant proteins were expressed in *E. coli* BL21 Δ*lacZ* (DE3)/pRARE2 and purified following standard protocols using His-affinity and size-exclusion chromatography. Mutants of *E. ramulus* HMPREF0373_02965 (*Er*Lnb136) were constructed using QuickChange II Site-Directed Mutagenesis (Agilent) with pETM11_ HMPREF0373_02965 as template. Primers used for site-directed mutagenesis are listed in Table S8 and mutants were produced as described above. L-Selenomethionine (L-SeMet) labelled protein expression of *Er*Lnb136 was performed by introducing the corresponding plasmid into *E. coli* B834 (DE3) and culturing the transformed cells in a synthetic M9 based medium of the SelenoMet labelling Kit (Molecular Dimensions) supplemented either with L-methionine or L-SeMet (both to 50 µg mL^{-1}). The L-SeMet labelled protein was purified as described above.

Growth experiments and single strain proteomics analysis.
R. hominis DSM 16839, *R. inulinivorans* DSM 16841, *E. ramulus* DSM 15684 and *E. ramulus* DSM 16296 were grown anaerobically at 37 °C using a Whitley DG250 Anaerobic Workstations (Don Whitley Scientific). *R. hominis* and *R. inulinivorans* were propagated in YCFA medium[41] while for *E. ramulus* strains CFA medium (modified YCFA medium lacking yeast extract to minimize *E. ramulus* growth on yeast extract) was used. Growth media were supplemented with 0.5% (w/v) carbohydrates sterilized by filtration (soluble carbohydrates, 0.45 µm filters) or autoclaving (mucins, 15 min at 121 °C) and cultures were performed in at least biological triplicates unless otherwise indicated. Bacterial growth was monitored by measuring $OD_{600 nm}$ and pH (for co-culture experiments). For growth experiments performed in microtiterplates, a Tecan Infinite F50 microplate reader (Tecan Group Ltd) located in the anaerobic workstation was used and growth was followed by measuring $OD_{595 nm}$. An unpaired two-tailed Student's *t*-test was used to determine the statistical significance between growth level reached between different culture conditions and non-carbohydrate controls.

For differential proteome analyses, *R. hominis* and *R. inulinivorans* were grown in 200 µL YCFA (1.5 mL Eppendorf tubes) to mid-late exponential phase (OD_{600} ~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×g, 5 min at 4 °C), washed twice with ice cold 0.9% (w/v) NaCl, resuspended in 20 µL lysis buffer (50 mM HEPES, 6 M GuHCl, 10 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP), 40 mM 2-chloroacetamide (CAA) pH 8.5) and stored at −80 °C for proteomics analysis.

Co-culture experiment and proteomics analyses.
R. hominis, *R. inulinivorans* and *A. muciniphila* DSM 22959 were grown in 10 mL YCFA to mid-late exponential phase (OD_{600} ~0.6-0.7). From these pre-cultures, equal amounts of cells (OD_{600}) were used to inoculate 30 mL fresh YCFA medium with 1% (w/v) of a mucin mixture (0.6% (w/w) 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 were performed in four biological replicates and growth was followed (OD_{600} and pH) at 0, 6, 8, 12, 16, 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 harvested by centrifugation (5000 × 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 until proteomics analysis. Collected 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[53,54]. 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 × 10 s at 4 °C) (Bioruptor, Diagenode). Lysates were centrifuged (14,000×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 1:3 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 1:1 with 2% (w/v) trifluoroacetic acid (TFA) to quench enzymatic activity and peptides were processed for mass spectrometry using in house packed stage tips[55] as described below.

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 µL of 80% (v/v) ACN, 0.1% (w/v) FA by centrifugation at 1800×g and equilibrated twice with 40 µL of 3% (v/v) ACN, 1% (w/v) FA before samples were loaded in steps of 50 µL. After loading, tips were washed three times with 100 µL 0.1% (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 200 µL tip and tips were activated and equilibrated by loading of 80 µL (v/v) 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 NH$_4$OAc, 20% (v/v) ACN, 0.5% (w/v) FA, then with 60 µL of 225 mM NH$_4$OAc, 20% (v/v) ACN, 0.5% (w/v) FA and lastly with 5% (v/v) NH$_4$OH, 80 % (v/v) ACN into 0.5 mL 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 trap column (Thermo Fisher 164705) using a Thermo EasyLc 1000 HPLC system. Peptides were eluted with a gradient of 4.8–48% (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. Full MS spectra were collected at 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 unknown species were excluded. MS/MS spectra were acquired at a resolution of 17,500, with an AGC target value 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 16839 (ID: UP000008178) or *R. inulinivorans* DSM 16841 (ID: UP000003561) respectively, as obtained from Uniprot. The spectra from co-culture 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 had to be identified in at least two out of the three samples analysed per culture condition with high confidence.

Relative bacterial abundance in co-cultures was estimated based on strain unique peptides identified with Unipept version 4.0[56]. 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 *Equate I and L* and *Advanced missed cleavage handling* activated. 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 H_2SO_4 for peak identification and quantification. Samples from four biological replicates were analysed by injecting 20 μL of standard or filtrated (0.45 μM filter) culture supernatant on a 7.8×300 mm Aminex HPX-87H column (Biorad) combined with a 4.6×30 mm Cation H guard column (Biorad). Elution of was performed with a constant flow rate of 0.6 mL min^{-1} and a mobile phase of 5 mM H_2SO_4. Standards were analysed as above in technical triplicates.

Oligosaccharide uptake preference of *Roseburia* spp. *R. hominis* was grown anaerobically in 250 μL YCFA medium with 0.5% (w/v) of an equal mixture of xylotetraose and LNT in biological triplicates. Samples (20 μL) 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 at 4 °C) before supernatants were stored at −20 °C until the HPAEC-PAD analysis. Standards of 0.5 mM xylotetraose and LNT were prepared in 100 mM NaOH and used to identify corresponding peaks in the chromatograms. Samples or standard were injected (2 μL injections) on a 4×250 mm CarboPac PA10 column with a 4×50 mm CarboPac guard column and eluted isocratically (0.750 mL min^{-1}, 100 mM NaOH, 10 mM NaOAc). The analysis was performed from a biological triplicate and standards were analysed in technical duplicates.

For determining uptake preference of Le^b tetraose, Le^a triose and blood group H triose type I, *R. inulinivorans* was grown anaerobically in 200 μL YCFA medium supplied with an equal mixture of 1.5 mM Le^b tetraose, 1.5 mM Le^a triose and 1.5 mM blood group H triose type I in biological triplicates. Samples (10 μL) were taken after 0, 3.5, 5.5, 6.5, 8, 9.5 and 24 h, diluted 10-fold in ice cold 20 mM NaOH and centrifuged (10 min at 5000×g at 4 °C) before supernatants were stored at −20 °C until the HPAEC-PAD analysis. Standards of 0.1 mM Le^b tetraose, Le^a triose and blood group H triose type I were prepared in 20 mM NaOH and used to identify corresponding peaks in the chromatograms. Samples or standard were injected (10 μL injections) on a 4×250 mm CarboPac PA10 column with a 4×50 mm CarboPac guard column and eluted isocratically (0.750 mL min^{-1}, 50 mM NaOH). The analysis was performed from a biological triplicate and standards were analysed in technical duplicates.

For investigating mixed HMO uptake, *R. inulinivorans* was grown anaerobically in 300 μL YCFA medium with 0.5% (w/v) of mixed HMOs purified from mother's milk or 0.5% (w/v) of mixed HMOs purified from mother's milk but previously digested with *Ri*Le$^{a/b}$136 (0.5 μM *Ri*Le$^{a/b}$136 for 18 h) in biological triplicates. Samples (10 μL) were taken after 0 and 24 h, diluted 10-fold in ice cold 20 mM NaOH and centrifuged (10 min at 5000×g at 4 °C) before supernatants were stored at −20 °C until the HPAEC-PAD analysis. Standards of 0.1 mM Le^b tetraose, Le^a triose, blood group H triose type I, LNDFH I, Lactose, 2′FL and LNT were prepared in 20 mM NaOH and used to identify corresponding peaks in the chromatograms. Samples or standard were injected (10 μL injections) on a 4×250 mm CarboPac

PA200 column with a 4×50 mm CarboPac guard column and eluted isocratically (0.350 mL min^{-1}, 50 mM NaOH). The analysis was performed from a biological triplicate and standards were analysed in technical duplicates.

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

Hydrolysis kinetics and specific activities of the GH136 lacto-*N*-biosidases were measured using a coupled enzymatic assay to monitor lactose release. The lactose was hydrolysed with a ß-galactosidase (used above) and the resulting glucose was oxidized with a glucose oxidase (Sigma Aldrich) concomitant with the production of H_2O_2 measured by coupling to horseradish peroxidase (Sigma Aldrich) oxidation of 4-aminoantipyrine and 3,5-dichloro-2-hydroxybensensulfonic acid. Reactions were prepared in 96-well microtiter plates to a final volume of 150 μL, containing substrate, lacto-*N*-biosidase, ß-galactosidase (150 U mL^{-1}), glucose oxidase (150 U mL^{-1}), horseradish peroxidase (150 U mL^{-1}), 10 mM 3,5-dichloro-2-hydroxybensensulfonic acid, 1 mM 4-aminoantipyrine in standard assay buffer. Reactions were performed at 37 °C and $A_{515 nM}$ was measured in 5 sec intervals for 30 min. Blanks were prepared by substituting lacto-*N*-biosidase with standard assay buffer in the reaction mixture and a lactose standard (3–500 μM) was used for the quantification.

Hydrolysis kinetics of *Rh*Lnb136 (40 nM) and *Er*Lnb136 (10 nM) towards LNT (0.2–5 mM for *Rh*Lnb136 and 0.1–2.5 mM for *Er*Lnb136) were determined as described above. The kinetic parameters K_M and k_{cat}, were calculated by fitting the Michaelis-Menten equation to the initial rate data using OriginPro 2018b and OriginPro 2019b (OriginLab). Lacto-*N*-biosidase specific activity of *Ri*Le$^{a/b}$136 (1.2 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.

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

Enzyme product profiles. Enzyme assays were performed at 37 °C for 16 h in standard assay buffer or in the phosphate version (instead of MES) for GH112 enzymes, in independent biological triplicates. Degradation products were analysed by thin layer chromatography (TLC) and or Matrix-assisted laser desorption/ionization time of flight mass 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[57]. 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 *Ri*Le$^{a/b}$136 was according to[58], following permethylation of oligosaccharides[59]. For permethylation, lyophilized oligosaccharides 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 % (w/v) acetic acid was added followed by the addition of 2 mL of CH_2Cl_2. 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-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 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 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 NaBH₄ 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[60]. 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)[61] were prepared as described above, except 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 NH₄HCO₃ 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 −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 analysed. The data were processed using Xcalibur software (version 2.0.7, Thermo Scientific).

Oligosaccharide binding analysis. Binding of LNT, LNB, GNB, H type I triose, Leᵃ triose and Leᵇ tetraose to *Ri*Leᵃ/ᵇBP was analyzed by surface plasmon resonance (SPR; Biacore T100, GE Healthcare). *Ri*Leᵃ/ᵇBP, diluted in 10 mM NaOAc buffer pH 3.75 to 50 μg mL⁻¹, was immobilized on a CM5 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 measured at 25 °C in 20 mM sodium phosphate buffer, 150 mM NaCl, 0.005% (v/v) P20 (GE Healthcare), pH 6.5. Experiments were performed in duplicates (each consisting of a technical duplicate) in the range of 0.3–50 μM for LNB, 0.78-200 μM for GNB, 0.97–250 μM for Leᵃ, 0.097–100 μM for Leᵇ and 1.5–250 μM for blood H type I triose. To investigate ligand specify of *Ri*Leᵃ/ᵇBP, binding was further tested towards 0.5 mM LNT, LN*n*T, lactose, blood A triose, 2′FL and 3′FL. Equilibrium dissociation constants (K_D) were calculated by fitting a one binding site model to steady state sensograms, using the Biacore T100 data evaluation software.

Binding of LNT, LNB, GNB, LNnT, lactose and 2′FL to *Rh*LNBBP was measured using a Microcal ITC₂₀₀ calorimeter (GE Healthcare). Titrations were performed in duplicates at 25 °C with *Rh*LNBBP (0.1 mM) in the sample cell and 1.5 mM ligand in 10 mM sodium phosphate buffer, pH 6.5 in the syringe. A first injection of 0.4 μL was followed by 19 injections of 2 μL ligand each, separated by 180 s. Heat of dilution was determined from buffer titrations and corrected data were analysed using MicroCal Origin software v7.0. To determine binding thermodynamics a non-linear single binding model was fitted to the normalized integrated binding isotherms.

Differential scanning calorimetry. The Differential scanning calorimetry (DSC) analyses was performed at protein concentrations of 1 mg mL⁻¹ in 20 mM sodium phosphate buffer, 150 mM NaCl, pH 6.5, using a Nano DSC (TA instruments). Thermograms were recorded from 10 to 90 °C at a scan speed of 1 °C min⁻¹ using buffer as reference. Baseline corrected data were analysed using the NanoAnalyze software (TA instruments). DSC analyses were performed in duplicates unless otherwise states.

Crystallization. Crystals of *Er*Lnb136 proteins were grown at 20 °C using the sitting-drop vapor diffusion method, by mixing 0.5 μL of a 10 mg mL⁻¹ protein solution with an equal volume of a reservoir solution. Native crystals were grown in a 20% (w/v) PEG4000, 0.1 M sodium citrate pH 5.6, and 20% isopropanol reservoir solution. SeMet-labelled crystals were grown using a reservoir solution containing 20% (w/v) PEG6000, 0.1 M Tris-HCl pH 8.5, and 1 M lithium chloride. The crystals were cryoprotected in the reservoir solution supplemented with 20% (v/v) glycerol and 25 mM LNB. The crystals were flash-cooled at 100 K (−173.15 °C) in a stream of nitrogen gas. Diffraction data were collected at 100 K on beamlines at SLS X06DA (Swiss Light Source, Swiss) and Photon Factory of the High Energy Accelerator Research Organization (KEK, Tsukuba, Japan). The data were processed using HKL2000[62] and XDS[63]. Initial phase calculation, phase improvement, and automated model building were performed using PHENIX[64]. Manual model rebuilding and refinement was achieved using Coot[65] and REFMAC5[66]. Because the crystal structures of SeMet-labelled and native protein were virtually the same (root mean square deviations for the Cα atoms = 0.14 Å), we used the SeMet-labelled protein structure for the descriptions in the Results and Discussion.

Molecular graphics were prepared using PyMOL (Schrödinger, LLC, New York) or UCSF Chimera (University of California, San Francisco)

Bioinformatics. SignalP v.4.1[67], PSORTb v3.0[68], TMHMM v.2.0[69] were used for prediction of signal peptides and transmembrane domains. InterPro[70] and dbCAN2[71] were used to analyse modular organization using default settings for Gram positive bacteria. Redundancy in biological sequence datasets was reduced using the CD-HIT server (sequence identity cut off = 0.95)[72]. Protein sequence alignments were performed using MAFFT (BLOSUM62)[73]. Phylogenetic trees were constructed using the MAFFT server, based on the neighbour-joining algorithm, and with bootstraps performed with 1000 replicates. Phylogenetic trees were visualized and tanglegrams constructed using dendroscope[74]. Colouring of protein structures according to amino acid sequence conservation was accomplished in UCSF Chimera, based on protein multiple (structural based) alignments from the PROMALS3D server[75] and by using the in UCSF Chimera implemented AL2CO algorithm[76]. The MEME suite web server was used for amino acid sequence motif discovery and evaluation[77]. Protein structures were compared using the Dali server (http://ekhidna2.biocenter.helsinki.fi/dali/) (PMID: 27131377) and the molecular interface between *Er*Lnb136₁ and *Er*Lnb136₁₁ was analysed (solvent inaccessible interface, Gibbs energy) via the PDBePISA server (https://www.ebi.ac.uk/pdbe/pisa/).

The abundance and distribution of HMO utilization genes encoding GH112, GH136₁ and GH136₁₁ in *Roseburia* were analysed by a BLAST search of the corresponding DNA reference sequences from *R. intestinalis* L1-82, *R. hominis* A2-183 and *R. inulinivorans* A2-194 against a total of 4599 reconstructed *Roseburia* genomes, binned into 42 Species-level Genome Bins (SGBs) by Pasolli et al.[33]. The variability of the *Roseburia* core xylanase (GH10) was determined similarly by blasting the DNA reference sequences from *R. intestinalis* L1-82 (ROSINTL182_06494) against the same dataset.

For further analyses, 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* SGBs, based on the assignment of Pasolli et al.[33]. The retrieved genomes were used to analyse the gene landscape around the GH112 gene. The RAST server[78] was used for gene annotation. Based on the annotation and coordinates of the genes, 11 genes upstream and downstream the GH112 were selected for gene landscapes 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 neighbourhood, considering the present or absent of the genes on the gene landscape and as well its position on the loci. We used the function *stringdistmatrix* with the "osa" 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 Reporting Summary 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 Supplementary Table 8. Data underlying Fig. 1a–d, Supplementary Figs. 2, 4c, 5a–f and 8a–c are provided as Source Data files and the reproducibility of representative experiments is indicated in the corresponding figure legends. All other data are available from the corresponding author upon requests. Source data are provided with this paper.

Code availability
The *stringdistmatrix* function is part of the R-project's *stringdist* package, available at GitHub (https://github.com/markvanderloo/stringdist) Source data are provided with this paper.

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**, 121–141 (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 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. 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**, 323–335 (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 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 colonic regulatory T cells. *Nature* **504**, 446–450 (2013).
20. Bultman, S. J. Bacterial butyrate prevents atherosclerosis. *Nat. Microbiol.* **3**, 1332–1333 (2018).
21. Canani, R. B. 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.* **294**, 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 Bifidobacterium longum. *Cell Chem. Biol.* **24**, 515–524.e5 (2017).
29. Leth, M. L. et al. Differential bacterial capture and transport preferences facilitate co-growth on dietary xylan in the human gut. *Nat. Microbiol.* **3**, 570–580 (2018).
30. Viborg, A. H. et al. Distinct substrate specificities of three glycoside hydrolase family 42 β-galactosidases from Bifidobacterium longum subsp. infantis ATCC 15697. *Glycobiology* **24**, 208–216 (2014).
31. Theilmann, M. C., Fredslund, F., Svensson, B., Lo Leggio, L. & Abou Hachem, M. Substrate preference of an ABC importer corresponds to selective growth on β-(1,6)-galactosides in Bifidobacterium animalis subsp. lactis. *J. Biol. Chem.* **294**, 11701–11711 (2019).
32. Derrien, M., Vaughan, E. E., Plugge, C. M. & de Vos, W. M. Akkermansia municiphila gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *Int. J. Syst. Evol. Microbiol.* **54**, 1469–1476 (2004).
33. Pasolli, E. et al. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. *Cell* **176**, 649–662.e20 (2019).
34. Wang, T. et al. Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. *ISME J.* **6**, 320–329 (2012).
35. Sakurama, H. et al. Lacto-N-biosidase encoded by a novel gene of Bifidobacterium longum subspecies longum shows unique substrate specificity and requires a designated chaperone for its active expression. *J. Biol. Chem.* **288**, 25194–25206 (2013).
36. Ejby, M. et al. Two binding proteins of the ABC transporter that confers growth of Bifidobacterium animalis subsp. lactis ATCC27673 on β-mannan possess distinct manno-oligosaccharide-binding profiles. *Mol. Microbiol.* **112**, 114–130 (2019).
37. Kostic, A. D. et al. The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. *Cell Host Microbe* **17**, 260–273 (2015).
38. Jost, T., Lacroix, C., Braegger, C. & Chassard, C. Assessment of bacterial diversity in breast milk using culture-dependent and culture-independent approaches. *Br. J. Nutr.* **110**, 1253–1262 (2013).
39. Collado, M. C., Delgado, S., Maldonado, A. & Rodríguez, J. M. Assessment of the bacterial diversity of breast milk of healthy women by quantitative real-time PCR. *Lett. Appl. Microbiol.* **48**, 523–528 (2009).
40. Jost, T., Lacroix, C., Braegger, C. P., Rochat, F. & Chassard, C. Vertical mother-neonate transfer of maternal gut bacteria via breastfeeding. *Environ. Microbiol.* **16**, 2891–2904 (2014).
41. Duncan, S. H., Hold, G. L., Barcenilla, A., Stewart, C. S. & Flint, H. J. Roseburia intestinalis sp. nov., a novel saccharolytic, butyrate-producing bacterium from human faeces. *Int. J. Syst. Evol. Microbiol.* **52**, 1615–1620 (2002).
42. La Rosa, S. L. et al. The human gut Firmicute Roseburia intestinalis is a primary degrader of dietary β-mannans. *Nat. Commun.* **10**, 905 (2019).
43. Scott, K. P., Martin, J. C., Duncan, S. H. & Flint, H. J. Prebiotic stimulation of human colonic butyrate-producing bacteria and bifidobacteria, in vitro. *FEMS Microbiol. Ecol.* **87**, 30–40 (2014).
44. Rossez, Y. et al. Almost all human gastric mucin O-glycans harbor blood group A, B or H antigens and are potential binding sites for Helicobacter pylori. *Glycobiology* **22**, 1193–1206 (2012).
45. Holgersson, J., Jovall, P. Å. & Breimer, M. E. Glycosphingolipids of human large intestine: detailed structural characterization with special reference to blood group compounds and bacterial receptor structures. *J. Biochem.* **110**, 120–131 (1991).
46. Anand, S., Kaur, H. & Mande, S. S. Comparative in silico analysis of butyrate production pathways in gut commensals and pathogens. *Front. Microbiol.* **7**, 1945 (2016).
47. Van Den Abbeele, P. et al. Butyrate-producing Clostridium cluster XIVa species specifically colonize mucins in an in vitro gut model. *ISME J.* **7**, 949–961 (2013).
48. Ward, R. E. Isolation of milk oligosaccharides using solid-phase extraction. *Methodology* **2**, 9–15 (2009).
49. Robinson, R. C., Colet, E., Tian, T., Poulsen, N. A. & Barile, D. An improved method for the purification of milk oligosaccharides by graphitised carbon-solid phase extraction. *Int. dairy J.* **80**, 62–68 (2018).
50. Miller, R. S. & Hoskins, L. C. Mucin degradation in human colon ecosystems: Fecal population densities of mucin-degrading bacteria estimated by a "most probable number" method. *Gastroenterology* **81**, 759–765 (1981).
51. Skoog, E. C. et al. Human gastric mucins differently regulate helicobacter pylori proliferation, gene expression and interactions with host cells. *PLoS ONE* **7**, e36378 (2012).
52. Dümmler, A., Lawrence, A.-M. & de Marco, A. Simplified screening for the detection of soluble fusion constructs expressed in E. coli using a modular set of vectors. *Microb. Cell Fact.* **4**, 34 (2005).
53. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* **11**, 319–324 (2014).
54. Ejby, M. et al. An ATP binding cassette transporter mediates the uptake of α-(1,6)-linked dietary oligosaccharides in bifidobacterium and correlates with competitive growth on these substrates. *J. Biol. Chem.* **291**, 20220–20231 (2016).
55. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896–1906 (2007).
56. Gurdeep Singh, R. et al. Unipept 4.0: functional analysis of metaproteome data. *J. Proteome Res.* **18**, 606–615 (2019).
57. Anderson, K., Li, S.-C. & Li, Y.-T. Diphenylamine–aniline–phosphoric acid reagent, a versatile spray reagent for revealing glycoconjugates on thin-layer chromatography plates. *Anal. Biochem.* **287**, 337–339 (2000).
58. Sugiyama, Y. et al. Introduction of H-antigens into oligosaccharides and sugar chains of glycoproteins using highly efficient 1,2-α-L-fucosynthase. *Glycobiology* **26**, 1235–1247 (2016).

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

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

61. 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).

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

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

64. 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).

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

66. Murshudov, 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).

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

68. 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).

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

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

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

72. 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).

73. 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).

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

75. 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).

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

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

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

Acknowledgements

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 *Ri*Le^a/b^136 was performed by A.G.,To.K., M.S. and Ta.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