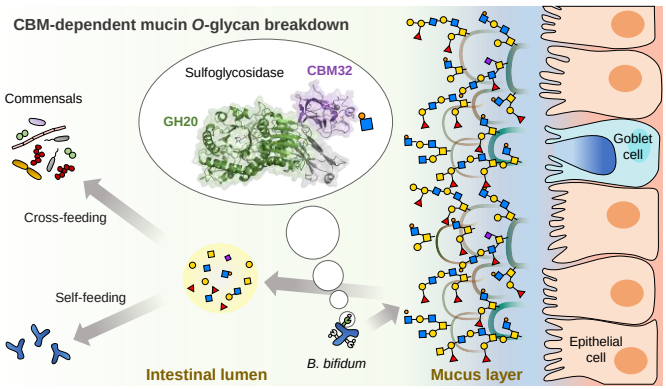

CBM-dependent mucin O-glycan breakdown
Sulfoglycosidase
CBM32
GH20
Commensals
Cross-feeding
Self-feeding
B. bifidum
Intestinal lumen
Mucus layer
Goblet cell
Epithelial cell

Title

A bacterial sulfoglycosidase highlights mucin *O*-glycan breakdown in the gut ecosystem

Author list

Toshihiko Katoh[1], Chihaya Yamada[2], Michael D. Wallace[3], Ayako Yoshida[2], Aina Gotoh[1], Moe Arai[1], Takako Maeshibu[1], Toma Kashima[2], Arno Hagenbeek[4], Miriam N. Ojima[1], Hiromi Takada[1], Mikiyasu Sakanaka[1], Hidenori Shimizu[1], Keita Nishiyama[5], Hisashi Ashida[6], Junko Hirose[7,#], Maria Suarez-Diez[4], Makoto Nishiyama[2], Ikuo Kimura[1], Keith A. Stubbs[3], Shinya Fushinobu[2]*, and Takane Katayama[1]*

Affiliations

[1]Graduate School of Biostudies, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan

[2]Grasuate School of Agricultural and Life Sciences, The University of Tokyo, Bunkyo-ku, Tokyo 113-8657

[3]School of Molecular Sciences, The University of Western Australia, 35 Stirling Hwy, Crawley, WA 6009, Australia

[4]Laboratory of Systems and Synthetic Biology, Wageningen University & Research, 6703 HB Wageningen, The Netherlands

[5]Department of Microbiology, School of Pharmacy, Kitasato University, Minato-ku, Tokyo 108-8641, Japan

[6]Faculty of Biology-Oriented Science and Technology, Kindai University, Kinokawa, Wakayama 649-6493, Japan

[7]School of Human Cultures, The University of Shiga Prefecture, Hikone, Shiga 522-8533, Japan

[#]Present address: Department of Food and Nutrition, Kyoto Women's University, Kyoto 605-8501, Japan

Corresponding authors: Takane Katayama (lead contact, takane@lif.kyoto-u.ac.jp) and Shinya Fushinobu (asfushi@mail.ecc.u-tokyo.ac.jp)

Abstract

Mucinolytic bacteria modulate host-microbiota symbiosis and dysbiosis through their ability to degrade mucin *O*-glycans. However, how and to what extent bacterial enzymes are involved in the breakdown process remains poorly understood. Here, we focus on a glycoside hydrolase family 20 sulfoglycosidase (BbhII) from *Bifidobacterium bifidum*, which releases *N*-acetylglucosamine-6-sulfate from sulfated mucins. Glycomic analysis showed that, in addition to sulfatases, sulfoglycosidases are involved in mucin *O*-glycan breakdown *in vivo* and that the released *N*-acetylglucosamine-6-sulfate potentially affects gut microbial metabolism, both of which were also supported by a metagenomic data mining analysis. Enzymatic and structural analysis of BbhII reveals the architecture underlying its specificity and the presence of a GlcNAc-6S-specific carbohydrate binding module (CBM) 32 with a distinct sugar recognition mode that *B. bifidum* takes advantage of to degrade mucin *O*-glycans. Comparative analysis of the genomes of prominent mucinolytic bacteria also highlights a CBM-dependent *O*-glycan breakdown strategy utilised by *B. bifidum*.

Main text

Introduction

The intestinal mucus layer is a frontline barrier that modulates gut microbe–host interactions. In the colon, MUC2, a highly *O*-glycosylated gel-forming mucin, is secreted primarily from goblet cells together with other proteins to form a mucus layer, which consists of a densely packed layer attached to the epithelium (attached layer) and a loosely packed outer layer. While the attached layer is essentially devoid of microbes, the outer layer harbors a dense microbial community[1]. A recent study showed that mucins secreted from the proximal colon encapsulate gut microbes to form stools[2], wherein mucin *O*-glycans, together with dietary fibers, serve as nutrients for bacteria[3,4]. Several *in vitro* studies show that mucin *O*-glycan mono- and oligosaccharides produced by mucinolytic bacteria are shared between microbial members to influence the metabolite outcome[5,6], indicating that bacterial mucin degradation is important for commensalism among different microbes and between microbes and the host. Mucin *O*-glycan degradation is related to both remission and progression of inflammatory bowel diseases[7] suggesting that mucinolytic bacteria serve as modulators of host-microbiota symbiosis and dysbiosis.

Mucin *O*-glycans are increasingly acidified from the proximal to the distal colon by sialylation and sulfation[2,8]. These modifications are thought to confer resistance to bacterial degradation of mucins, but several mucinolytic bacteria exploit sialidases[9] and sulfatases[10–12] to remove such modifications. Accordingly, these decapping enzymes have been targets of study as they permit further bacterial mucin breakdown[12]. However, there could be another unexplored pathway for the sulfated *O*-glycan degradation. Glycoside hydrolase (GH) family 20 sulfoglycosidases[13], which were first isolated from *Prevotella* sp. (Sgl)[14] and then from *Bifidobacterium bifidum* (BbhII)[15], can release *N*-

acetylglucosamine-6-sulfate (GlcNAc-6S) from porcine gastric mucin (PGM) *in vitro*. This activity does not seem to be the result of *N*-acetylglucosaminidase promiscuity, as BbhII showed 400-fold higher activity toward GlcNAc-6S over GlcNAc residues[15].

B. bifidum is a Gram-positive anaerobe capable of assimilating host glycans such as human milk oligosaccharides and mucin *O*-glycans but is incapable of plant polysaccharide degradation[16,17]. This bacterium possesses cell surface-anchored GHs acting on almost all glycosidic linkages in *O*-glycans, with the known exception of α-linked *N*-acetylgalactosaminides[16]. As a possible reflection of this repertoire of GHs, *B. bifidum* colonises the intestines of a wide range of mammals[18]. However, its *in vivo* mucin *O*-glycan degradative capability has not been addressed and is controversial[19]. Here, through structural, glycomic, and informatics studies on BbhII combined with animal and human sample analyses, we not only demonstrate the *in vivo* relevance of sulfoglycosidase to intestinal mucin *O*-glycan breakdown but also reveal the mechanistic basis of how *B. bifidum* takes advantage of a novel GlcNAc-6S-specific carbohydrate-binding module (CBM) 32[13], found within BbhII to degrade *O*-glycans. Comparative genomic analysis of mucinolytic microbes highlights a CBM-dependent mucin *O*-glycan strategy employed by *B. bifidum*.

Results

B. bifidum affects intestinal *O*-glycan metabolism in mice

Using conventional mice we examined how *B. bifidum* administration affects intestinal mucin *O*-glycan breakdown (Fig. 1a). Quantitative PCR analysis indicated, at day 5, *B. bifidum*-colonization of the caecum contents and faeces, but not of the intestinal surface, at an abundance of 0.062% and 3.7% per total 16S rRNA genes, respectively (Supplementary Table 1). The data were comparable with faecal microbiota composition analysis (Extended Data Fig. 1a). Using faecal extracts containing soluble mucins collected (day 5), semi-quantitative *O*-glycomic analysis was performed (Fig. 1b,c and Supplementary Table 2). Estimated total amounts of non-sulfated and sulfated *O*-glycan species were comparable between the control (PBS) and *B. bifidum*-administered groups (Fig. 1d); however, when we compared the ratio (%) of each glycan (*B. bifidum*/control group) as a function of oligosaccharide length, a negative correlation was detected (Fig. 1e). Longer *O*-glycan oligosaccharides decreased with a corresponding increase in shorter oligosaccharides upon *B. bifium*-administration.

We quantified mucin *O*-glycan-constituting monosaccharides, *i.e.*, L-fucose (Fuc), galactose (Gal), *N*-acetylgalactosamine (GalNAc), *N*-acetylneuraminic acid (NeuAc), *N*-acetylglucosamine (GlcNAc), and GlcNAc-6S in caecum contents (Fig. 1f and Supplementary Fig. 1a) and faeces (Fig. 1g) at day 5. Sulfated Gal was not considered, as no bacterial GHs have been annotated as a sulfo-galactosidase. GlcNAc-6S abundance, relative to total monosaccharides, in the caecum contents was higher in the *B. bifidum*-administered group than in the control group (Fig. 1f). The faecal GlcNAc-6S concentration was not different between the two groups at day 5. However, within the *B. bifidum*-

administered group, the concentration decreased in faecal samples at day 5 compared to day 0, while no such change was observed between day 0 and day 5 control group samples (Fig. 1g). The results suggested that caecum GlcNAc-6S release triggered by *B. bifidum*-administration shifts faecal microbiota metabolism to assimilate the monosaccharide. Analysis using the Linear discriminant analysis Effect Size (LEfSe) algorithm of faecal microbiota revealed an increase of the genus *Bacteroides* only in the day 5 samples from the *B. bifidum*-administered group (Extended Data Fig. 1b). Blastn analysis revealed that the amplicon sequence variants (ASVs) correspond to *Phocaeicola sartorii*, *Bacteroides oleiciplenus*, and *Bacteroides rodentium.* The former two species possess GlcNAc-6S sulfatase homologues (SulfAtlas S1_11)[20] of *Bacteroides thetaiotaomicron* VPI-5482[12] (>70% protein identity with BT_3177). Prevalence of the homolog in available genomes of *P. sartorii* and *B. oleiciplenus* are 71% (5/7) and 60% (3/5), respectively. Fresh mouse faecal suspensions indeed consumed exogenously added GlcNAc-6S, whilst heat-treated samples did not (Supplementary Fig. 1b). Overall, these results indicate the involvement of sulfoglycosidases in mucin *O*-glycan degradation *in vivo* and suggests an impact of released GlcNAc-6S on faecal microbiota.

Human sample and metagenomic data-mining analyses

We quantified free GlcNAc-6S in faeces of human infants and adults. GlcNAc-6S amounts were lower in adult compared to infant samples (Fig. 1h, right panel). A positive correlation was observed between the amounts of free GlcNAc-6S and the *bbhII* gene in infant samples, whereas in adult samples no correlation was observed (Fig. 1h, left panel). The range of *bbhII* abundance was similar between the two groups. We hypothesized that the lower GlcNAc-6S concentration seen in adult samples might be due to GlcNAc-6S consumption by other microbes, as seen in mouse samples. Human faecal suspensions incubated in the presence of 10 mM GlcNAc-6S resulted in a similar microbiota shift. Weighted UniFrac distance analysis did not show a marked difference between control and GlcNAc-6S-added groups post-cultivation; however, the increase of the genus *Bacteroides* was detected by LEfSe with a linear discriminant analysis (LDA) score of > 4 (Extended Data Fig. 2). Blastn analysis showed that the ASVs correspond to *Bacteroides stercoris* and *Bacteroides finegoldii*, whose genomes encode GlcNAc-6S sulfatase homologues with 71% (50/82) and 79% (22/28) prevalence, respectively.

We mined a human feacal metagenomic dataset[21] to examine potential correlations between abundances of *bbhII* and bacterial taxon therein. *sgl*, another characterised sulfoglycosidase gene from *Prevotella* sp.[14], was also included. We first constructed a phylogenetic tree of characterised GH20 members (Extended Data Fig. 3), also including BbhII and Sgl homologues (> 40% in amino acid identity). The results showed that BbhII and Sgl homologues form distinct clades. Sgl and its homologue BT_4394 share the same node with other GH20 members with different specificities, *e.g.* chitinolytic activity[22], therefore we considered these two homologues only. Metagenomic data analysed included the reads of 80 mother-preweaning infant pairs at 4 months post-delivery (1.67×10^9

± 3.61×10^6 reads/sample). A positive correlation between the abundances of BbhII clade genes and *Flintibacter* sp. was detected in the mother's samples, in addition to *B. bifidum* (Fig. 1i and Extended Data Fig. 4a). The genome of *Flintibacter* sp. KGMB00164, belonging to the phylum *Bacillota*, does not encode a sulfoglycosidase but encodes a sulfatase-like homologue (F3I61_RS00920). The modeled structure of F3I61_RS00920 overlapped with that of a GlcNAc-6S sulfatase (BT_3177, PDB ID: 7P24)[23] at the active site (Supplementary Fig. 2). Regarding Sgl homologues, positive correlations were detected with many *Bacteroides* spp. in both mother and infant samples (Extended Data Fig. 4b,c). This can be attributed to the co-occurrence of sulfatase and sulfoglycosidase genes in the genus, as among 189 completed genomes, all 76 Sgl-positive strains possess GlcNAc-6S-sulfatase homologues.

Faecal mucin *O*-glycan breakdown in a gnotobiotic mouse model

To examine the mucinolytic activity of *B. bifidum* in detail, whilst ruling out any indirect effects of other microbes, we established a *B. bifidum*-mono-colonised mice. A *bbhII* mutant was also used in the experiment (Extended Data Fig. 5a,b). The wild-type (WT) and *bbhII* strains were administered to germ-free mice once at day 0, while PBS was administered to the control group, with the caecum and faeces in the colon collected at day 5 (Fig. 2a). qPCR analysis of caecum contents showed that both strains colonised intestines for the experiment duration with a lower colonization observed for *bbhII* (Fig. 2b). Free monosaccharides in caecum contents were detected in higher amounts for WT-colonised than for *bbhII*-colonised mice although some of them were not statistically significant, while in the germ-free mice they were not detected or negligible (Fig. 2c). While the NeuAc concentration was proportional to the abundance of *B. bifidum* cells in caecum contents within the three groups, the GlcNAc-6S amounts was high only in the WT group compared to *bbhII*- and control groups (Fig. 2d, e). The GlcNAc-6S concentration was comparable between *bbhII*- and control groups, demonstrating the role of BbhII in host glycan degradation. *O*-Glycomic analysis of faecal extracts provided evidence of the ability of *B. bifidum* to breakdown mucin *O*-glycans and the role of BbhII in the process (Fig. 2f-l and Supplementary Table 3). In the WT group, the estimated total amounts of non-sulfated and sulfated *O*-glycans in faecal extracts decreased to 6 and 10%, respectively, compared to the control group (Fig. 2i,j). A similar reduction was also observed for the *bbhII* group, but the amount of remaining sulfated *O*-glycans was higher in the *bbhII* group than in the WT group. No significant difference was detected for the non-sulfated glycan amounts between the two groups. Examination of the individual sulfated *O*-glycan species revealed that a peak at *m*/*z* 867 remained unconsumed in the *bbhII* group compared to the WT group (Fig. 2k). MS/MS analysis predicted the glycan to be a GlcNAc-sulfated mucin core 2 structure (Fig. 2l). We found that *B. bifidum* degrades keratan sulfate among several glycosaminoglycans (GAGs) (Extended Data Fig. 6), and therefore the small amounts of Gal and GlcNAc-6S may be derived from this GAG.

BbhII-susceptible and resistant mucin *O*-glycans *in vitro*

To determine the substrate specificity of BbhII, *O*-glycomes of porcine colonic mucin (PCM) were analysed pre- and post-incubation with a purified enzyme (WTc-His_6, Supplementary Fig. 3a). A Volcano plot showed a decrease of eight sulfated *O*-glycan species upon BbhII digestion among the total 59 sulfated and non-sulfated *O*-glycans identified (Supplementary Table 4). The deduced glycosyl composition of the most abundant BbhII-susceptible glycan was $(SO_3^-)_1Fuc_1Gal_1HexNAc_1GalNAc$-itol (*m/z* 1041.4 $[M+2Na-H]^+$). MS/MS fragments generated from the precursor ion predicted a terminal GlcNAc-sulfate-containing, fucosylated core 2 structure (Extended Data Fig. 7). The other seven BbhII-susceptible *O*-glycans also gave a diagnostic peak of the sulfated core 2 structure (*m/z* 631). BbhII-resistant sulfated *O*-glycans (*e.g. m*/*z* 1245.5) were deduced to contain an internal GlcNAc-sulfate moiety (Extended Data Fig. 8). Partially decreased peaks (*e.g. m*/*z* 1490.7) likely represent mixtures of BbhII-susceptible and resistant structures. BbhII thus acts on the terminal 6-*O*-sulfated GlcNAc moiety.

Structural and mutational analyses of BbhII

To elucidate the structural basis that differentiates sulfoglycosidases from β-*N*-acetylhexosaminidases, we solved the crystal structure of BbhII. The protein comprises a signal peptide [1−32 amino acid residues (aa)], carbohydrate-binding module 32 (CBM32, 46−197 aa), a GH family 20b domain (GH20b, 205−344 aa), a GH20 catalytic domain (347−714 aa), and a C-terminal transmembrane region (1,029−1,044 aa) (Supplemental Fig. S3). Deletion analysis revealed that a construct comprising 39−861 aa retained both stability and activity comparable to the full-length BbhII (Supplemental Fig. S3b).

The structure of BbhII WTc (39−861 aa)-His_6 complexed with GlcNAc-6S was solved at 1.65 Å resolution (Fig. 3a and Supplementary Table 5) and revealed three domains: an N-terminal β-sandwich domain assigned as CBM32 (46−197 aa, CBM32 N-domain), a GH20 catalytic domain with a $(\beta/\alpha)_8$-barrel fold (198−750 aa), and a C-terminal β-sandwich domain (751−861 aa, C-domain). Two GlcNAc-6S molecules are bound in the CBM32 N-domain and in the GH20 catalytic domain. A Dali search[24] revealed that BbhII resembles the β-hexosaminidase from *Streptomyces plicatus* (*Sp*HEX, PDB ID: 1HP5)[25] with the root-mean-square deviation (RMSD) = 2.0 Å for 473 Cα atoms (Z score = 31.6). The closest homolog of the BbhII CBM32 N-domain was the chitosan-binding module DD2 of *Paenibacillus* sp. chitosanase/glucanase (PDB ID: 4ZZ5)[26], with RMSD = 1.5 Å for 128 Cα atoms (Z score = 20.2). The C-domain shares structural similarity with immunoglobulins (PDB ID: 4EZM; Z score = 7.7 with RMSD = 3.3 Å for 87 Cα atoms).

β-GlcNAc-6S binds to the subsite (−1) with a ^4E conformation (Supplementary Fig. 4a). The carbonyl oxygen of *N*-acetyl group is positioned 2.7 Å from the C1 atom of the sugar, reflecting

substrate-assisted catalysis adopted by GH20 members[27]. The pseudo-axial O1 atom and the acid/base residue Glu553 forms a hydrogen bond, and the *N*-acetyl group is hydrogen-bonded with Asp552 (the stabilizer) and Tyr637 (Fig. 3b). An aromatic cage formed by Trp588, Trp607, and Trp685 stabilizes the distorted conformation of the *N*-acetyl group, and Trp685 also forms a stacking interaction with the sugar ring. The O3 and O4 atoms of the sugar are recognised by the side chains of Arg358 and Glu687, while the sulfate group interacts with Gln640 and Trp651 by direct hydrogen bonding and with His688 by water-mediated hydrogen-bonds. Unexpectedly, the hydrophobic side chains of Pro639 and Val649 surround this area and no basic residues in this region are involved in the sulfate group recognition. Superimposition with GlcNAc-thiazoline-complexed *Sp*HEX[25] showed conservation of the catalytically important residues (Asp552/Asp553/Tyr637) whilst revealing differences around the C6-position between the enzymes (Fig. 3c). In BbhII, Trp651, corresponding to Trp408 of *Sp*HEX, is shifted away to accommodate the sulfate group and BbhII substitutes Pro639, Gln640, Val649, and His688 for Asp395, Met396, Leu406, and Thr455 of *Sp*HEX. These replacements might confer a specific sulfoglycosidase activity to BbhII, as the enzyme shows 400-fold higher catalytic efficiency (k_{cat}/K_m) for *p*-nitrophenyl (*p*NP)-β-GlcNAc-6S over *p*NP-β-GlcNAc[15]. The above mentioned 12 residues, except for Val649, are conserved within the BbhII clade of the GH20 tree, while in the two neighboring homologues in the tree (CAB72127.1 and AAQ05800.1) P639 and Q640 that interact with the sulfate group are replaced with more bulky aspartate and methionine residues, respectively (Extended Data Fig. 3). CAB72127.1 has been identified to be a chitinase[28]. The sulfoglycosidase activity could thus be acquired by widening the substrate pocket of a β-hexosaminidase towards C6 position of GlcNAc in the clade.

Alanine scanning mutations demonstrated the importance of Asp552 (stabilizer) and Glu553 (acid/base) in catalysis (Extended Data Fig. 5d and Supplementary Table 6). The specific activity of the mutants decreased >1000-fold compared to WTc-His_6. Alanine-replacement of Pro639, Gln640, and Val649 slightly increased the activity, while W651A had slightly decreased activity. When the two hydrogen bonds between the sulfate group and BbhII were disrupted by Q640A/W651A, the activity decreased (~60%). The higher k_{cat}/K_m value of BbhII for *p*NP-GlcNAc-6S over *p*NP-GlcNAc is primarily attributed to the 90-fold smaller K_m value[15], indicating that the sulfate group contributes to the tight active site binding. Trp651, assisted by Gln640, could play a role in the recognition of 6-sulfate group. There is no steric hindrance for the binding of *galacto* sugars, which accords with the ability of BbhII to hydrolyse *p*NP-β-GalNAc-6S as effectively as *p*NP-β-GlcNAc-6S[15].

Characterization of the synthesised BbhII inhibitors

We synthesised two putative sulfoglycosidase inhibitors, *O*-(2-acetamido-2-deoxy-6-*O*-sulfo-D-glucopyranosylidene)amino *N*-phenylcarbamate (PUGNAc-6S, Supplementary Fig. 5) and 1,2-dideoxy-2′-methyl-α-D-glucopyranoso-[2,1-d]-Δ2′-thiazoline-6-sulfate (NAGT-6S) based on known

hexosaminidase inhibitors[29,30] and examined their efficacy using BbhII . The inhibition observed for both compounds was competitive with K_i values of 15.4 and 52.3 nM for PUGNAc-6S and NAGT-6S, respectively (Fig. 4a and Extended Data Fig. 9a,b) with these data in accord with those of their parental β-hexosaminidase inhibitors: K_i = 40 nM for *Monilesaurus rouxii* enzyme (PUGNAc)[31] and 280 nM for Jack Bean β-HexNAcase (NAGT)[30].

A crystal structure of PUGNAc-6S-complexed BbhII was determined at 2.23 Å resolution (Supplementary Table 5). Clear electron densities for two PUGNAc-6S molecules were observed in the CBM32 N-domain and the GH20 catalytic site (Fig. 4b,c). The RMSD between GlcNAc-6S- and PUGNAc-6S-complexed structures is 0.223 Å for 734 Cα atoms. The PUGNAc-6S in the active site adopts a ^4E conformation similar to that of GlcNAc-6S. The interactions at the catalytic site is also similar between the two with the key hydrogen bonds with the acid/base catalyst (Glu553) and the stabilizer (Asp552) retained. Although the electron density map of the PUGNAc-6S phenyl group is ambiguous, a stacking interaction between the moiety and Trp651 might confer the tighter binding potency of PUGNAc-6S over NAGT-6S.

The effects of NAGT-6S on microbial mucin breakdown were examined. We used this inhibitor due to its superior stability over PUGNAc-6S in the medium. *B. bifidum* was cultivated in PGM-supplemented media in the absence and presence of NAGT-6S, which was followed by *O*-glycomic analysis (Fig. 4d and Supplementary Table 7). In the absence of NAGT-6S, the estimated total amounts of both non-sulfated and sulfated *O*-glycans decreased to 25% post-cultivation (Fig. 4e,f). In the presence of NAGT-6S, inhibition was observed for sulfated *O*-glycan degradation in a dose-dependent manner, and a sulfated core 2 *O*-glycan (*m*/*z* 867) accumulated markedly. The MS/MS fragmentation pattern of the peak was consistent with that of the peak which remained unconsumed in *bbhII*-administered mice (Fig.2l and Supplementary Fig. 6). The results indicate that further degradation of the *m*/*z* 867 *O*-glycan, which is formed by defucosylation of a sulfated glycan at *m*/*z* 1041 by 1,2-α-L-fucosidase[32], likely requires BbhII-mediated prior removal of GlcNAc-6S. This inhibitor is thus a useful tool for understanding enzyme structure and function and in predicting how bacteria sequentially degrade natural substrates *in vivo*.

Role of the CBM32 in mucin *O*-glycan breakdown by *B. bifidum*

In the CBM32 N-domain, both GlcNAc-6S and PUGNAc-6S adopted a $^4C_1$-like conformation (Supplementary Fig. 4b). The relatively high θ value (21.6°) of PUGNAc-6S may be due to the sp^2-hybridised character of the C1 atom (Fig. 4a). GlcNAc-6S forms a stacking interaction with Trp183 and directly hydrogen bonds with Glu62, Asn89, Arg95, and Asn126 (Fig. 5a). Ser97, Thr127, and Ser184 also form water-mediated hydrogen bonds with the sugar. Ca^{2+} is coordinated near the GlcNAc-6S binding site but is not directly involved in carbohydrate-binding. A similar Ca^{2+} binding site is found in other CBM32 structures[33]. For PUGNAc-6S the *N*-phenylcarbamate is solvent-exposed,

making its electron density ambiguous (Fig. 4b) thereby suggesting that BbhII CBM32 specifically binds to terminally attached GlcNAc-6S and does not capture internal GlcNAc-6S residues. Glu62, Asn89, Arg95, Ser97, Asn126, and Trp183 are conserved in the BbhII homologues, with the exception that Trp183 is substituted for tyrosine or phenylalanine in some orthologues. The BbhII CBM32 and its closest structural homolog, a chitosan-binding module DD2 of *Paenibacillus* sp. chitosanase/glucanase[26] were compared (Fig. 5b). In contrast to the single-sided recognition of the ligand by DD2, the BbhII CBM32 recognises GlcNAc-6S from both sides of the sugar ring with an additional short helix in the variable loop region. A GH84 exo-β-*N*-acetylglucosaminidase from *Clostridium perfringens* has an unusual CBM32 (NagH CBM32-2) specific for a terminal GlcNAc[33], but its binding site is shallower than the GlcNAc-6S binding site of BbhII CBM32 (Fig. 5b). Although CBM32 is known as one of the most diversified families of CBMs[13] with an assorted repertoire of ligand specificities and recognition architectures, the BbhII CBM32 is a novel example in this family.

To analyse the sugar-binding specificity of BbhII CBM32, we employed an enzyme-linked immune solvent assay (ELISA). A CBM-His_6 construct (39-200 aa: Supplementary Fig. 3a) was used. Several trials showed that binding is detectable when glycoproteins were fixed on the plate and probed with CBM-His_6 pre-complexed with Penta-His mouse IgG and anti-mouse IgG-HRP. The binding plots obtained for PGM and PCM were fitted to a typical saturation curve (Fig. 5c). Higher titers observed for PCM than for PGM indicated higher amounts of terminal GlcNAc-6S in PCM *O*-glycans. At the saturating concentration of PCM, inhibitory effects of sugars were examined (Fig. 5d). The binding was abolished by either GlcNAc-6S or *p*NP-β-GlcNAc-6S, diminished moderately by *p*NP-β-GlcNAc-3S, *p*NP-β-GlcNAc-3,4diS and, to a lesser extent, by *p*NP-β-GalNAc-6S. These results, together with structural insights, indicate a high specificity of BbhII CBM32 for terminal GlcNAc-6S residues of *O*-glycans, but further research using sulfated GAG oligosaccharides is needed. The BbhII CBM32 is likely specific for *gluco*-configured sugars as the side chain of Asn89 appears to sterically encumber the axial O4 of *galacto*-configured sugars (Fig. 5a).

The thermodynamics of CBM-His_6 binding to *p*NP-β-GlcNAc-6S was analysed using isothermal titration calorimetry (Extended Data Figs. 9c,d). Stoichiometric binding occurs (1:1) and proceeds through an enthalpy-driven process with an unfavorable entropy change. The affinity (K_d = 25 μM) is relatively high amongst other CBM32s with fairly weak binding sites[26,34,35]. No heat pulse was detected during the *p*NP-β-GlcNAc titration nor during the titration of CBM-His_6(W183A) with *p*NP-β-GlcNAc-6S, demonstrating the importance of the sulfate group of the sugar and the role of Trp183 in binding.

To explore the biological relevance of this CBM32, we expressed three BbhII variants in *Bifidobacterium longum* JCM 31944, which is sulfoglycosidase-negative. Plasmids carrying the entire *bbhII* gene (WT), the W183A substituted gene (W183A), or the CBM32-deleted gene (ΔCBM) were constructed (Supplementary Fig. 3a). Western blot using anti-BbhII antibodies confirmed similar

levels of expression of three variants in *B. longum* (Extended Data Fig. 5f). Recombinant enzymes with respective mutations were also prepared (Extended Data Fig. 5g). *p*NP-β-GlcNAc-6S-hydrolysing activity was comparable among the three variants regardless of whether purified preparations or recombinant *B. longum* cells were used for the assay (1~1.6-fold change, Fig. 5e). However, when PGM was used as the substrate, the GlcNAc-6S releasing ability of the purified enzymes was reduced by 2.2-fold and 4.3-fold by W183A-His_6 and ΔCBM-His_6, respectively, as compared with WTc-His_6 (Fig. 5f). Remarkably, the recombinant cells expressing W183A and ΔCBM showed 2.5-fold and 10-fold lower GlcNAc-6S releasing activity than WT-expressing cells, respectively. Thus, *B. bifidum* benefits from this novel GlcNAc-6S-specific CBM32 to efficiently attach to and degrade mucin *O*-glycans.

Richness of CBMs and loss of endo-*O*-glycanase in *B. bifidum*

Given the important role of the CBM32 in mucin *O*-glycan breakdown, we compared the prevalence and abundance of CBMs among prominent mucinolytic gut microbes. CBM32, 40, 47, 51, and 71 were extracted as possible mucin *O*-glycan-related (*muc*-) CBMs from the genomes of *B. bifidum*, *Akkermansia muciniphila*, *Bacteroides caccae*, *Bacteroides fragilis*, *Bacteroides thetaiotaomicron*, *B. bifidum*, *Clostridium perfringens*, *Prevotella melaninogenica*, and *Ruminococcus gnavus* (Supplementary Table 8). We analysed how the distribution of these *muc*-CBMs is associated with the occurrence of possible *muc*-GHs in each bacterial species/strain, by using Bray-Curtis distance analysis followed by permutational multivariate analysis of variance (PERMANOVA) (Extended Data Fig. 10). Consequently, the highest effect size was detected for GH16 subfamily 3 (GH16_3)[36] that comprises endo-*O*-glycanases that release oligosaccharides from mucin *O*-glycans[37]. Further analyses revealed that GH16_3 is conserved in all strains belonging to *A. muciniphila* (except one), three *Bacteroides* species, and *P. melaninogenica*, while it is not highly conserved in *C. perfringens* (9/46) and is absent in *R. gnavus*. In the *B. bifidum* genomes (8/11), GH16_3 was present as a loss-of-function gene (Fig. 6a) and was absent in the remaining strains. Further analysis revealed that GH16_3-negative mucinolytic bacteria such as *B. bifidum* and 37 strains of *C. perfringens* possess significantly higher numbers of possible *muc*-CBMs embedded in *muc*-GHs in the same polypeptides and a higher ratio of possible *muc*-GHs with *muc*-CBMs per total *muc*-GHs, than GH16_3-positive bacteria including 9 strains of *C. perfringens* (Fig. 6b and Supplementary Fig. 7). *R. gnavus* was an exception which has a specialized genetic toolset for sialic acid assimilation[38]. These results suggest CBM-dependent and GH16_3-dependent *O*-glycan breakdown strategies in gut microbes.

Discussion

The genomes of mucinolytic bacteria encode a variety of *muc*-GHs, and animal experiments have shown that some of these GHs are upregulated when dietary fiber is limited[3,4]. A recent study revealed

that sulfatases, which enhance the accessibility of GHs to *O*-glycans by decapping the terminal and internal sulfate groups, play a crucial role for *Bacteroides* species to competitively colonise mouse intestines[12]. These results indicate that mucins form the environment for certain microbes to initiate and establish interaction with their host. However, how and to what extent mucinolytic bacteria decompose mucin *O*-glycans using *muc*-GHs *in vivo* and thereby affecting the gut microbial community remains largely unknown. In the present study we showed that *B. bifidum* degrades mucin *O*-glycan *in vivo* (Figs. 1 and 2) by focusing on a sulfoglycosidase that releases GlcNAc-6S from sulfated *O*-glycans. The saccharide concentration was positively correlated with the amount of *bbhII* in infant stools, although no correlation was detected for adult samples. This can be rationalized by infant gut microbiota being generally rich in bifidobacteria[21], while adult gut microbiota is generally rich in *Bacteroides* species, many of which possess both sulfatase and sulfoglycosidase. Incubation of GlcNAc-6S with mice and human adult faecal suspensions increased the abundance of *Bacteroides* sp., which was accompanied with GlcNAc-6S consumption (Extended Figs. 1 and 2). On the other hand, in the metagenomic dataset of adults, *bbhII* abundance was positively correlated with the abundance of *Flintibacter* sp. which harbors a sulfatase gene (Fig. 1i). Thus, GlcNAc-6S residues in *O*-glycans are decomposed either by sulfoglycosidase-mediated release followed by desulfation to GlcNAc or sulfatase-mediated decapping followed by release of GlcNAc by a β-hexosaminidase, in which both inter- and intraspecies cooperation is involved. The polysaccharide utilization locus for the latter scenario was identified in *Bacteroides* genomes[39], while *B. bifidum* employs the first pathway and mediates cross-feeding since it can't assimilate GlcNAc-6S. The *bbhII* gene does not form a cluster with other related genes in *B. bifidum* genomes, but it is a member of the *nagR* regulon consisting of the genes dedicated to host glycan degradation[40].

The BbhII CBM32 seems to be highly specific for terminal GlcNAc-6S residues (Figs. 4 and 5), which is unprecedented among the listed CBM32s in CAZy[34], and its binding mode is distinct from known CBM32s[26,33–35]. This unique CBM32 enhanced the GlcNAc-6S releasing capability of transformed *B. longum* cells (Fig. 5f). Localization of the CBM32 on the bacterial surface enables the cells to efficiently capture sulfated *O*-glycans on glycoproteins, which potentially increases their fitness in competitive environments. This is in sharp contrast to the sulfoglycosidases of *Prevotella* sp. Sgl and *B. thetaiotaomicron* BT_4394, neither of which have known CBMs.

Lack of a GH16_3 endo-*O*-glycanase was associated with both the high abundance of possible *muc*-CBMs in *muc*-GHs and the high ratio of possible *muc*-CBM-carrying *muc*-GHs per total *muc*-GHs, with the exception of *R. gnavus* (Fig. 6b). Notably, the same trends were observed even within a single species (*C. perfringens*). It should be mentioned, however, that our analysis included some GH members that also target carbohydrates other than mucin *O*-glycans and that *Bacteroides* spp. and other Gram-negative bacteria are shown to have sugar-binding domains unclassified or unrelated to known CBMs[41]. Regarding *B. bifidum*, it is unclear whether the pseudogenisation of GH16_3 has

evolved *muc*-CBM-carrying *muc*-GHs or if the emergence of *muc*-CBM-carrying *muc*-GHs compromised the pre-existing GH16_3. Nonetheless, the dependence of *B. bifidum* on surface-localised CBMs for efficiently degrading mucin *O*-glycans with cognate GHs becomes apparent (Fig. 6c). Cross-feeding to other bacteria can also occur because *B. bifidum* leaves Fuc, Gal, NeuAc, and GlcNAc-6S unconsumed[42]. It is interesting to note that *Bifidobacterium breve* possesses a gene set for GlcNAc-6S utilization[43]. A GH16_3 negative, CBM scarce *R. gnavus* has a different system in which a *trans*-sialidase releases 2,7-anhydro-Neu5Ac from *O*-glycans, for the utilization of which a specific transporter is required[44].

Overall, our findings warrant further research to elucidate how gut microbes benefit from CBMs, endo-*O*-glycanases, or other yet unidentified glycan-binding domains and GHs to persist in the gut ecosystem and how these pathways influence microbiome formation in the gut ecosystem. As such, *B. bifidum*, with its large repertoire of GHs and CBMs specifically acting on host glycans, is a promising organism to explore the involvement of these modules in mucin decomposition in the gut, an environment rich in both plant and host glycans.

Acknowledgements:

This study was partly supported by JSPS-KAKENHI (17K17820 and 19K05789 to TosK, 21H02116 to TaK), a grant-in-aid from the Institution for Fermentation, Osaka (2015 to TosK), and a JSPS Research Fellowship (17J08530 to AG and 21J15883 to HT). We thank Takatoshi Arakawa (The University of Tokyo) for helping X-ray data collection and inhibition kinetics assay, staff of the Photon Factory and SPring-8 (proposal No. 2019B2556) for X-ray data collection, Saeko Nagao (Nagao Midwife Clinic, Kyoto, Japan) for helping infant stool collection, Fumihiko Sato, Kentaro Ifuku, and Takeshi Nakano (Kyoto University) for MALDI-TOF/MS instrument maintenance, Jun Wada (Kyoto Integrated Science and Technology Bio-Analysis Center) for LC-MS/MS technical assistance, Masanori Yamaguchi (Wakayama University) for providing *p*NP-β-GlcNAc-3S and *p*NP-β-GlcNAc-3,4diS, and Akira Yoshimi (Kyoto University) for technical assistance on sugar analysis. MDW and KAS thank the facilities, and the scientific and technical assistance of Microscopy Australia at the Centre for Microscopy, Characterisation and Analysis, The University of Western Australia, a facility funded by the University, State and Federal Commonwealth Governments. MDW is supported by a Research Training Program Scholarship provided by the Australian Federal Government and The University of Western Australia.

Author Contributions Statement

TosK and TaK conceived the project and designed the experiments. TosK, TM, and MA performed glycomic analysis and enzyme characterization. AG, TosK, MA, MNO, HS, HT, IK, and TaK

conducted animal experiments and microbiota analysis. HT performed monosaccharide analysis. JH collected infant samples and managed the metadata. HT and MS constructed a *bbhII* mutant of *B. bifidum*. CY, TomK, and SF determined the protein structures and the inhibition constants. AY and MN are responsible for ITC analysis. HA constructed the full-length BbhII expression plasmid. MDW and KAS synthesised inhibitors. KN prepared PCM. AH, MSD, and TosK performed metagenomic data mining analysis. TosK, KAS, SF, and TaK drafted and edited the manuscript. All authors discussed the data and contributed to the completion of the manuscript.

Competing Interests Statement

Employment of MNO and MS at Kyoto University is in part supported by Morinaga Milk Industry Co., Ltd. Employment of HS at Kyoto University is supported by Noster Inc. The authors declare no other conflicts of interest.

441 ***Figure Legends***

442 **Figure 1. GlcNAc-6S release in mouse and human intestines. *a*−*g*,** *B. bifidum* (*Bb*) was
443 administered to conventional mice (C57BL/6N). ***a***, Experimental plan. ***b***, Representative MALDI-
444 TOF/MS profiles of permethylated *O*-glycans of faecal extracts of a *Bb*-administered mouse (Day 5).
445 External standards (red), lacto-*N*-fucopentaose I* (LNFP I) and sulfo-Lewis[a] trisaccharide**, for non-
446 sulfated and sulfated glycans, respectively. ***c***, Heatmap showing the estimated amounts of *O*-glycan
447 species obtained from PBS- and *Bb* groups. Means of a single technical replicate (STR) for each
448 sample (n = 5/group) were plotted. ***d***, Comparison of *O*-glycan amounts between PBS- and *Bb* groups
449 at Day 5 (two-tailed Welch's *t*-test). The bars and whiskers represent mean ± standard deviation (SD).
450 ***e***, The *Bb*/PBS group ratio of each *O*-glycan amount as a function of oligosaccharide length with non-
451 sulfated (black circle) and sulfated *O*-glycans (white circle) represented. All plots were used for two-
452 tailed Spearman's rank correlation analysis. ***f***, Relative abundance (%) of each of mucin *O*-glycan-
453 constituting monosaccharides per total monosaccharides in caecum contents was compared between
454 the two groups (two-tailed Welch's *t*-test, n = 5/group). The mean ± SD of a STR for each sample are
455 shown by bars and whiskers. ***g***, Comparison of GlcNAc-6S amounts in faeces between Day 0 and Day
456 5 within the PBS group (left) or the *Bb* group (right) (two-tailed paired *t*-test, n = 5/group). Values of
457 a STR for each sample were plotted. ***h* and *i,*** *In vivo* and *in silico* human sample analyses. ***h***, Two-
458 tailed Spearman's rank correlation analysis between the abundances of *bbhII* and free GlcNAc-6S in
459 infant (n = 33, purple) and adult (n = 18, green) faeces (left). GlcNAc-6S amounts were also compared
460 between infant and adult groups (right). Boxes represent median with interquartile range, while
461 whiskers represent the minimum and maximum variations (two-tailed Mann-Whitney test). *bbhII* was
462 quantified by PCR in duplicate and the means are reported. GlcNAc-6S was quantified from a STR
463 for each sample. ***i***, Bacterial species whose abundance shows a significant positive correlation with
464 *bbhII* homolog abundance in a metagenomic dataset[21]. The data of eighty mother-unweaning infant
465 pairs at 4 months post-delivery were used for two-tailed Spearman's rank correlation analysis.
466

467 **Figure 2. Intestinal mucin *O*-glycan breakdown in *B. bifidum*-mono-colonised mice. *a***,
468 Experimental plan. *B. bifidum* wild-type (WT) or *bbhII* mutant cells were administered to germ-free
469 mice (ICR, n = 8/WT and *bbhII* groups) at Day 0. The control group received PBS (n = 7). At Day 5,
470 mice were euthanized. ***b***, *B. bifidum* in caecum contents were quantified by PCR in duplicate and the
471 median with interquartile range (MIR) of each sample are presented (n = 7/control; n = 8/WT and
472 *bbhII* groups). A dashed line indicates the detection limit. ***c***, Comparison of mucin *O*-glycan-
473 constituting monosaccharide concentrations among the groups. The MIR of a STR for each sample
474 are presented (n = 7/control; n = 8/WT and *bbhII* groups). A Kruskal-Wallis test followed by Dunn's
475 multiple comparison test was used. ***d* and *e***, Free NeuAc (***d***) and GlcNAc-6S (***e***) concentrations as a
476 function of *B. bifidum* cells in caecum contests from the three groups. A positive correlation was

observed between Neu5Ac and *B. bifidum* amounts (two-tailed Spearman's rank correlation analysis). The GlcNAc-6S concentration was statistically higher in the WT group than in the other two groups (***c*** and ***e***). ***f* and *g***, Full-mass spectra of permethylated non-sulfated (***f***) and sulfated (***g***) *O*-glycans of faecal extracts from the three groups (Day 5). External standards (red), as in Fig. 1. ***h***, Heatmap showing the estimated amounts of *O*-glycan species of faecal extracts from the three groups. The upper limit is set to 0.2 nmol/0.1 mg faecal extracts. The mean of a STR for each sample (n = 7/control and WT groups; n = 8/*bbhII* group) were used for plotting. ***i*–*k***, Estimated amounts of total non-sulfated (***i***) and sulfated (***j***) *O*-glycans and of an ion peak of m/z 867 (***k***) from faecal extracts of the three groups (n = 7/control and WT groups; n = 8/*bbhII* group). The bars and whiskers represent MIR of the respective groups. A Kruskal-Wallis test followed by Dunn's multiple comparison test was used. ***l***, A MS/MS spectrum at m/z 867 obtained from a *bbhII*-mono-colonised sample. A deduced structure is shown with its fragmentation patterns with glycan symbols depicted (inset).

Figure 3. Structural analysis of BbhII. ***a***, Overall structure of BbhII (WTc-His_6: 39-861 aa) complexed with two GlcNAc-6S molecules shown as a ribbon representation solved at 1.65 Å resolution. Two GlcNAc-6S molecules (yellow sticks) bind at the CBM32 N-domain and the GH20 catalytic domain. The Ca^{2+} ion bound to the CBM32 is shown as a green sphere. ***b***, GlcNAc-6S in the GH20 catalytic domain (see Supplementary Fig. 4a for the sugar conformation). The sulfate group is recognised by direct and water-mediated hydrogen bonds. The amino acid residues and waters involved in the binding of GlcNAc-6S are shown with green sticks and red spheres, respectively. ***c***, Comparison of the catalytic sites of BbhII (green) and *Streptomyces plicatus* β-*N*-acetylhexosaminidase (*Sp*Hex1, cyan) complexed with GlcNAc-thiazoline (magenta) (PDB ID: 1HP5)[25].

Figure 4. Characterization of synthesized competitive inhibitors of BbhII. ***a***, Structures of PUGNAc-6S and NAGT-6S. K_i values were calculated by non-linear regression of the S-v plots obtained for WT-His_6 BbhII in the absence and presence of the inhibitors using the substrate *p*NP-β-GlcNAc-6S (Extended Data Fig. 9a,b). ***b* and *c***, Enlarged view of PUGNAc-6S bound to the CBM32 N-domain (***b***) and the GH20 catalytic domain (***c***) of BbhII (WTc-His_6) solved at 2.23 Å resolution. Coloring is the same as in Fig. 3, except that the carbon atoms of PUGNAc-6S are shown in dark green. Polder maps (σ = 3.0) are shown in blue. Surrounding residues, including those interacting via hydrogen bonds or stacking are represented in licorice. ***d*–*f,*** NAGT-6S-mediated inhibition of PGM *O*-glycan degradation in bacterial culture. *B. bifidum* was grown in basal medium supplemented with 1% PGM in the absence and presence of NAGT-6S (0.1 and 1 mM) for 24 h. ***d***, MALDI-TOF/MS profiles of permethylated *O*-glycans released by β-elimination from PGM collected prior (top panel) and post-cultivation (bottom two panels). External standards (red), as in Fig. 1. ***e* and *f***, Amounts of

non-sulfated (***e***) and sulfated (***f***) *O*-glycans estimated from MS spectra obtained in ***d***. Deduced glycan structures of *m*/*z* 867 and 1041 are shown with glycan symbols depicted (inset).

Figure 5. A novel CBM32 specific for GlcNAc-6S is pivotal for BbhII to efficiently release GlcNAc-6S from mucin *O*-glycans. ***a***, Binding mode of GlcNAc-6S (yellow) in the CBM32 (green) of BbhII WTc-His_6. ***b***, Comparison of the BbhII CBM32 with the chitosan-specific CBM32 (magenta) of chitosanase/glucanase (DD2) from *Paenibacillus* sp. in complex with chitotrisaccharide (cyan) (PDB ID: 4ZZ8) and the GlcNAc-specific CBM32 (orange) of GH84 exo-β-*N*-acetylglucosaminidase from *Clostridium perfringens* (NagH CBM32-2) in complex with GlcNAcβ1-2Man (cyan) (PDB ID: 2WDB). ***c***, Binding of the BbhII CBM32 to PGM (purple) and PCM (green), assessed by ELISA. The assay was conducted with (solid lines) and without (dashed lines) CBM-His_6. Mean ± SD of triplicate assays are shown by circles and whiskers, respectively with plots fitted to the normal saturation curve. ***d***, Inhibitory effect of various sugars (0.25 mM) on the CBM32-His_6 binding to PCM (10 μg/mL), assessed by ELISA. The bars and whiskers represent mean ± SD of triplicate assays. The values represent percentages of the control mean. Ordinary ANOVA followed by two-tailed Dunnett's test was used for statistical analysis. P values of < 0.05 are indicated. ***e*** **and** ***f***, The ability of purified BbhII variants and *B. longum* cells expressing BbhII variants to release GlcNAc-6S from *p*NP-β-GlcNAc-6S (***e***) and PGM (***f***). The variants include wild-type BbhII (WT), W183A mutant (W183A), and CBM32-deletion mutant (ΔCBM). Data are the mean ± SD of three independent assays, represented by the bars and whiskers (two-tailed Dunnett's test).

Figure 6. A possible link between the richness of mucin *O*-glycan-related CBMs and the functionality of endo-*O*-glycanase in mucinolytic gut microbes. ***a***, Phylogenetic tree of GH16 subfamily 3 (GH16_3) of selected mucinolytic bacteria. The fragmented non-functional sequences from *B. bifidum* are also included in the analysis. Amino acid sequences and their locus_tags were retrieved from the CAZy database and analysed by Clustal Omega with the tree constructed with FigTree v1.4.4 (http://tree.bio.ed.ac.uk/). Branch colors show the respective species; blue, *B. bifidum*; yellow, *Clostridium perfringens*; pink, *Akkermansia muciniphila*; green, *Bacteroides caccae*; purple, *Bacteroides fragilis*; orange, *Bacteroides thetaiotaomicron*; khaki, *Prevotella melaninogenica*; and gray, *Ruminococcus gnavus*. ***b***, The number of possible mucin *O*-glycan-related (*muc*-) CBMs in *muc*-GHs (upper panel) and the ratio of possible *muc*-GHs with *muc*-CBM per total *muc*-GHs (lower panel) in the selected mucinolytic bacterial genomes. Data are represented by dots with MIR. The numbers indicate the genomes examined ($n = 11, 37, 9, 92, 5, 19, 10, 12$, and 4 for *B. bifidum*, GH16_3-negative *C. perfringens*, GH16_3-positive *C. perfringens*, *A. muciniphila*, *B. caccae*, *B. fragilis*, *B. thetaiotaomicron*, *P. melaninogenica*, and *R. gnavus*, respectively), while −/+ indicates the absence or presence of intact GH16_3 in the genomes. Note that *C. perfringens* is divided into GH16_3-positive

549 and negative groups, depending on strains. A Kruskal-Wallis test followed by Dunn's multiple
550 comparison test was used for comparison with different letters indicating statistical significances (see
551 Supplementary Fig. 7 for P values). ***c***, Schematic representation of the possible CBM-dependent mucin
552 *O*-glycan breakdown strategy adopted by *B. bifidum*. Mucins are mainly produced from goblet cells
553 of intestinal epithelia. *B. bifidum*, and probably GH16_3-negative *C. perfringens* strains as well,
554 interact with mucin *O*-glycans via cell surface-located CBMs to initiate *O*-glycan breakdown using
555 cognate *muc*-GHs. The released mono- and disaccharides are utilized by *B. bifidum* and are also cross-
556 fed to other gut microbes which likely affects microbial diversity in the gut ecosystem. Glycan symbols
557 are shown in inset.

Extended Data Figure legends

Extended Data Figure 1. 16S rRNA gene-based mouse faecal microbiota analysis. ***a***, Relative abundances of bacterial taxa at the family level. Faeces of PBS- and *Bb*-administered conventional mice (n = 5/group) at Day 0 and Day 5 were used for the microbiota analysis. ***b***, The LDA score was calculated using LEfSe algorithm. Relative abundances of > 0.1% were used for the analysis.

Extended Data Figure 2. 16S rRNA gene-based microbiota analysis of human faecal suspensions incubated in the absence and presence of GlcNAc-6S. ***a,*** Relative abundances of bacterial taxa at the family level. Faecal samples obtained from 5 individuals were used for cultivation. Microbiotas were analysed pre- and post 24 h cultivation in the absence (none-added) and presence (GlcNAc-6S-added) of 10 mM GlcNAc-6S. ***b,*** Analysis of β-diversity among the samples, based on weighted UniFrac distance metrics. ***c***, The LEfSe analysis at the species level comparing between microbiotas of the none-added and GlcNAc-6S-added faecal suspensions post 24 h incubation. Relative abundances of > 0.1% were used for the analysis.

Extended Data Figure 3. Phylogenetic tree constructed using characterised GH20 members and sulfoglycosidase homologues. Amino acid sequences of characterised GH20 enzymes (CAZy database)[13], uncharacterised BbhII homologues of >40% identity (WP_172192827.1, WP_153878949.1, CRH87835.1, WP_206666329.1, WP_125968884.1, and WP_076060111.1), and an uncharacterised Sgl homologue of > 40% identity (ADJ68333.1, AAA65915.1, AAC44672.1, BAF76001.1, and AAO75563.1) were aligned by clustal omega with the tree constructed with FigTree v1.4.4. BbhII and Sgl clades are indicated. The reported substrate specificities for the homologues are indicated by circles with different colors. The sources of the homologues are also shown by different colors. The sequences classified into the BbhII and Sgl clades were used for analysing a deposited human metagenomic dataset[21] (Fig. 1*i* and Extended Data Fig. 4).

Extended Data Figure 4. Correlation analysis between the abundance of sulfoglycosidase (BbhII or Sgl)-specific reads and the abundance of each bacterial species-specific reads in a deposited metagenomic dataset. Bacterial species whose abundances show statistically significant correlations (q < 0.05) with the abundance of *bbhII* homologues (***a***), *sgl* homologues (***b***, ***c***), in the metagenomic dataset[21] are shown. Relative abundances of species- and gene-specific reads were calculated as described in the Methods section and used for two-tailed Spearman's rank correlation analysis. The reads of eighty mother (***a, b***)-unweaning infant (***c***) pair samples at 4 months post-delivery were used for the analysis.

Extended Data Figure 5. Confirmation of *bbhII* disruption in *B. bifidum*, heterologous

expression of BbhII in *B. longum*, and recombinant protein preparation. ***a***, Schematic representation of the *bbhII* gene inactivation by a single crossover recombination event. Primers used for the construction of a suicide vector (Pr-MS955 and Pr-MS956) are shown (Supplementary Table 9). The numbers of *B. bifidum* cells in mouse intestines and the *bbhII* gene in human faeces were determined by qPCR with a primer pair of bbhIIrt-P2-F and bbhIIrt-P2-R (Fig. 1*h*, Fig. 2*b*, and Supplementary Table 1). ***b***, Western blot analysis examining the expression of BbhII in *B. bifidum*. The cell-free extracts prepared from *B. bifidum* WT and *bbhII* mutant cells were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), followed by the detection using anti-BbhII antibodies. As a loading control, the expression of galacto-*N*-biose/lacto-*N*-biose I-binding protein (GLBP) was detected with anti-GLBP antibodies[45]. The images obtained in a single experiment are shown. These two strains were used for mono-colonisation of germ-free mice. ***c***, GlcNAc-6S-releasing activity of cell-free extracts prepared from *B. bifidum* WT and *bbhII* mutant cells. *p*NP-β-GlcNAc-6S was used at the concentration of 2 mM. Data are mean ± SD of three independent assays, represented by the bars and whiskers. ***d and e***, The results of SDS-PAGE of purified BbhII variants used for GlcNAc-6S-releasing assay (***d***) and ELISA and ITC analysis (***e***). The images of the gels obtained in a single experiment are shown. ***f,*** Heterologous expression of BbhII. Cell-free extracts prepared from recombinant *B. longum* strains harboring BbhII variant genes on plasmids (WT, W183A, and ΔCBM, Supplementary Fig. 3a) were separated by SDS-PAGE, followed by the detection using anti-BbhII antibodies. GLBP was used as the loading control. A representative image obtained in duplicate experiments is shown with essentially the same results obtained. The recombinant cells were used for examining the GlcNAc-6S-releasing activity from PGM *O*-glycans. ***g,*** The results of SDS-PAGE of purified BbhII-His_6 variants (WT, W183A, and ΔCBM, Supplementary Fig. 3a). The image of the gel obtained in a single experiment is shown.

Extended Data Figure 6. Degradation of glycosaminoglycans by *B. bifidum*. Heparan sulfate (HS), keratan sulfate (KS), chondroitin sulfate A (CS), and hyaluronan (HA) (0.4% each) were incubated with *B. bifidum* cell suspensions (equivalent to OD_{600} = 0.4) for 24 h at 37 ºC, and the reaction mixtures were analysed by thin-layer chromatography. The data obtained in a single experiment are shown. PGM was used as a positive control. Standard sugars used are Fuc, GlcNAc, GalNAc, Gal, GlcNAc-6S, and NeuNAc.

Extended Data Figure 7. Identification of BbhII-susceptible and resistant *O*-glycan structures of PCM. PCM was incubated in the absence and presence of purified BbhII (WTc-His_6). *O*-Glycans were then analysed with MALDI-TOF/MS. ***a***, The representative full-mass profiles (*m/z* 400-2400) of non-sulfated (left) and sulfated (right), permethylated *O*-glycan fractions obtained from non-treated (upper panels) and BbhII-treated (lower panels) PCM. The ion peaks shown in red are the externally added

standards: *LNFP I (*m/z* 1100.5) and **sulfo-Lewis^a trisaccharide (*m/z* 780.4). The analysis was conducted in technical triplicate. ***b***, A Volcano plot comparing non-treated with BbhII-treated PCM *O*-glycans. Fold-changes of the estimated glycan amounts and their *q*-values were plotted. The *q*-values are the adjusted *p*-values obtained by multiple *t* tests with false discovery rate correction with $Q = 5\%$ with the mean ± SD of three independent experiments used for evaluation. ***c***, A MS/MS spectrum of the most abundant, BbhII-susceptible sulfated glycan at *m/z* 1041 obtained in ***a***. The deduced *O*-glycan structure is shown with its fragmentation pattern. Glycan symbols are shown in inset.

Extended Data Figure 8. BbhII-resistant *O*-glycan structures of PCM. *a* and *b*, Data obtained in Extended Data Fig. 7 were further analysed here. ***a***, A MS/MS spectrum of the BbhII-resistant peak at *m/z* 1246 obtained from non-treated PCM. Two predicted glycan structures are shown with their fragmentation patterns. ***b***, MS/MS spectra of the *m/z* 1491 peaks obtained from non-treated PCM (upper panel) and BbhII-treated PCM (lower panel). The proposed BbhII-susceptible and resistant glycan structures are shown with their fragmentation patterns. The peaks and *m*/*z* values of the characteristic fragment ions formed from a predicted BbhII-susceptible glycan are shown in red (upper MS/MS profile). These peaks were not formed when the *m/z* 1491 peak obtained from BbhII-treated sample was subjected to MS/MS analysis (lower MS/MS profile). Glycan symbols are shown in inset.

Extended Data Figure 9. Biochemical analyses of BbhII. *a* and *b,* Inhibition of BbhII-catalysed reaction by the synthesised inhibitors. S-*v* plots (left panels) and Lineweaver-Burk plots (right panels) of *p*NP-β-GlcNAc-6S hydrolysis by BbhII WTc-His$_6$ in the absence and presence of PUGNAc-6S (***a***) and NAGT-6S (***b***). Inhibitor concentrations are shown in the insets. The kinetic parameters were calculated by curve-fitting experimental data to the Michaelis-Menten equation with competitive inhibition with the equation used for fitting shown. The results obtained from a single experiment were used for calculating the parameters. ***c* and *d,*** ITC analysis of BbhII CBM32 N-domain. Thermograms and binding isotherms obtained for *p*NP-β-GlcNAc-6S (left) and *p*NP-β-GlcNAc (right) are shown in the top and bottom panels, respectively. WT (***c***) and W183A (***d***) CBM-His$_6$ were used for the analysis. The concentrations and *c*-value are shown in the insets. Values of association constant (K_a), enthalpy of binding (ΔH), and binding stoichiometry (n) are expressed with the standard errors from a single fit to one set of sites model. Dissociation constants (K_d) were calculated from the reciprocal of K_a. The Gibbs free energy change (ΔG^0) and the entropy change (ΔS^0) were calculated from the equations $\Delta G^0 = -RT\ln K_a$ and $T\Delta S^0 = \Delta H - \Delta G^0$, respectively ($R$, gas constant; T, absolute temperature). The results obtained from a single technical replicate were used for calculating the parameters.

Extended Data Figure 10. Possible association between the abundance of *muc*-CBMs and the prevalence of *muc*-GHs in the prominent mucinolytic gut microbes. Exploratory analysis

examining effect size and significance of presence and absence of possible *muc*-GHs on the distribution of possible *muc*-CBMs in the genomes of selected mucinolytic bacterial species was performed NMDS, followed by a PERMANOVA with 9,999 iterations. NMDS on the distribution of *muc*-CBMs was used for ordination based on Bray-Curtis distances. R^2 and P values are shown in the table. The colors are*:* blue, *B. bifidum*; yellow, *Clostridium perfringens*; pink, *Akkermansia muciniphila*; green, *Bacteroides caccae*; purple, *Bacteroides fragilis*; orange, *Bacteroides thetaiotaomicron*; khaki, *Prevotella melaninogenica*; and gray, *Ruminococcus gnavus*.

References

1. Johansson, M. E. v. *et al.* The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. *Proc Natl Acad Sci U S A* **105**, 15064–15069 (2008).
2. Bergstrom, K. *et al.* Proximal colon-derived *O*-glycosylated mucus encapsulates and modulates the microbiota. *Science* **370**, 467–472 (2020).
3. Desai, M. S. *et al.* A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. *Cell* **167**, 1339-1353.e21 (2016).
4. Pudlo, N. A. *et al.* Symbiotic human gut bacteria with variable metabolic priorities for host mucosal glycans. *mBio* **6**, 1–12 (2015).
5. Pichler, M. J. *et al.* Butyrate producing colonic Clostridiales metabolise human milk oligosaccharides and cross feed on mucin via conserved pathways. *Nat Commun* **11**, 3285 (2020).
6. Belzer, C. *et al.* Microbial metabolic networks at the mucus layer lead to diet-independent butyrate and vitamin B12 production by intestinal symbionts. *mBio* **8**, e00770–17 (2017).
7. Yamada, T. *et al.* Mucin *O*-glycans facilitate symbiosynthesis to maintain gut immune homeostasis. *EBioMedicine* **48**, 513–525 (2019).
8. Robbe, C., Capon, C., Coddeville, B. & Michalski, J.-C. Structural diversity and specific distribution of *O*-glycans in normal human mucins along the intestinal tract. *Biochem J* **384**, 307–316 (2004).
9. Tailford, L. E., Crost, E. H., Kavanaugh, D. & Juge, N. Mucin glycan foraging in the human gut microbiome. *Front Genet* **6**, 81 (2015).
10. Wright, D. P., Knight, C. G., Parkar, S. G., Christie, D. L. & Roberton, A. M. Cloning of a mucin-desulfating sulfatase gene from *Prevotella* strain RS2 and its expression using a *Bacteroides* recombinant system. *J Bacteriol* **182**, 3002–3007 (2000).
11. Praharaj, A. B., Dehury, B., Mahapatra, N., Kar, S. K. & Behera, S. K. Molecular dynamics insights into the structure, function, and substrate binding mechanism of mucin desulfating sulfatase of gut microbe *Bacteroides fragilis*. *J Cell Biochem* **119**, 3618–3631 (2018).
12. Luis, A. S. *et al.* A single sulfatase is required to access colonic mucin by a gut bacterium. *Nature* **598**, 332–337 (2021).
13. Drula, E. *et al.* The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res* **50**, D571–D577 (2022).
14. Rho, J. *et al.* A novel mechanism for desulfation of mucin: identification and cloning of a mucin-desulfating glycosidase (sulfoglycosidase) from *Prevotella* strain RS2. *J Bacteriol* **187**, 1543–51 (2005).

15. Katoh, T. *et al.* Identification and characterization of a sulfoglycosidase from *Bifidobacterium bifidum* implicated in mucin glycan utilization. *Biosci Biotechnol Biochem* **81**, 2018–2027 (2017).
16. Katoh, T. *et al.* Enzymatic adaptation of *Bifidobacterium bifidum* to host glycans, viewed from glycoside hydrolyases and carbohydrate-binding modules. *Microorganisms* **8**, 481 (2020).
17. Turroni, F. *et al.* Genome analysis of *Bifidobacterium bifidum* PRL2010 reveals metabolic pathways for host-derived glycan foraging. *Proc Natl Acad Sci U S A* **107**, 19514–19519 (2010).
18. Milani, C. *et al.* Unveiling bifidobacterial biogeography across the mammalian branch of the tree of life. *ISME J* **11**, 2834–2847 (2017).
19. Kim, M. J. *et al.* Safety Evaluations of *Bifidobacterium bifidum* BGN4 and *Bifidobacterium longum* BORI. *Int J Mol Sci* **19**, (2018).
20. Barbeyron, T. *et al.* Matching the diversity of sulfated biomolecules: creation of a classification database for sulfatases reflecting their substrate specificity. *PLoS One* **11**, e0164846- (2016).
21. Bäckhed, F. *et al.* Dynamics and stabilization of the human gut microbiome in the first year of life. *Cell Host Microbe* **17**, 690–703 (2015).
22. Howard, B. M., Ekborg, A. N., Taylor, E. L., Weiner, M. R. & Hutcheson, W. S. Genomic analysis and initial characterization of the chitinolytic system of *Microbulbifer degradans* strain 2-40. *J Bacteriol* **185**, 3352–3360 (2003).
23. Luis, A. S. *et al.* Sulfated glycan recognition by carbohydrate sulfatases of the human gut microbiota. *Nat Chem Biol* **18**, 841–849 (2022).
24. Holm, L. DALI and the persistence of protein shape. *Protein Sci* **29**, 128–140 (2020).
25. Mark, B. L. *et al.* Crystallographic evidence for substrate-assisted catalysis in a bacterial β-hexosaminidase. *J Biol Chem* **276**, 10330–10337 (2001).
26. Shinya, S. *et al.* Mechanism of chitosan recognition by CBM32 carbohydrate-binding modules from a *Paenibacillus* sp. IK-5 chitosanase/glucanase. *Biochem J* **473**, 1085–1095 (2016).
27. Greig, I. R., Zahariev, F. & Withers, S. G. Elucidating the nature of the *Streptomyces plicatus* β-hexosaminidase-bound intermediate using ab initio molecular dynamics simulations. *J Am Chem Soc* **130**, 17620–17628 (2008).
28. Lonhienne, T. *et al.* Cloning, sequences, and characterization of two chitinase genes from the antarctic *Arthrobacter* sp. strain TAD20: isolation and partial characterization of the enzymes. *J Bacteriol* **183**, 1773–1779 (2001).
29. Knapp, S. *et al.* NAG-thiazoline, an *N*-acetyl-β-hexosaminidase inhibitor that

implicates acetamido participation. *J Am Chem Soc* **118**, 6804–6805 (1996).

30. Beer, D., Maloisel, J.-L., Rast, D. M. & Vasella, A. Synthesis of 2-acetamido-2-deoxy-D-gluconhydroximolactone- and chitobionhydroximolactone-derived N-phenylcarbamates, potential inhibitors of β-*N*-acetylglucosaminidase. *Helv Chim Acta* **73**, 1918–1922 (1990).
31. Horsch, M., Hoesch, L., Vasella, A. & Rast, D. M. *N*-Acetylglucosaminono-1,5-lactone oxime and the corresponding (phenylcarbamoyl)oxime. Novel and potent inhibitors of β-*N*-acetylglucosaminidase. *Eur J Biochem* **197**, 815–818 (1991).
32. Katayama, T. *et al.* Molecular cloning and characterization of *Bifidobacterium bifidum* 1,2-α-L-fucosidase (AfcA), a novel inverting glycosidase (glycoside hydrolase Family 95). *J Bacteriol* **186**, 4885–4893 (2004).
33. Ficko-Blean, E. & Boraston, A. B. N-Acetylglucosamine recognition by a family 32 carbohydrate-binding module from Clostridium perfringens NagH. *J Mol Biol* **390**, 208–220 (2009).
34. Abbott, D. W., Eirín-López, J. M. & Boraston, A. B. Insight into ligand diversity and novel biological roles for family 32 carbohydrate-binding modules. *Mol Biol Evol* **25**, 155–167 (2008).
35. Teh, A.-H., Sim, P.-F. & Hisano, T. Structural basis for binding uronic acids by family 32 carbohydrate-binding modules. *Biochem Biophys Res Commun* **533**, 257–261 (2020).
36. Viborg, A. H. *et al.* A subfamily roadmap of the evolutionarily diverse glycoside hydrolase family 16 (GH16). *J Biol Chem* **294**, 15973–15986 (2019).
37. Crouch, L. I. *et al.* Prominent members of the human gut microbiota express endo-acting O-glycanases to initiate mucin breakdown. *Nat Commun* **11**, 4017 (2020).
38. Tailford, L. E. *et al.* Discovery of intramolecular trans-sialidases in human gut microbiota suggests novel mechanisms of mucosal adaptation. *Nat Commun* **6**, 7624 (2015).
39. Martens, E. C., Chiang, H. C. & Gordon, J. I. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* **4**, 447–457 (2008).
40. Arzamasov, A. *et al.* Human milk oligosaccharide utilization in intestinal bifidobacteria is governed by global transcriptional regulator NagR. *mSystems* **7**, e00343-22 (2022).
41. Terrapon, N. *et al.* PULDB: the expanded database of Polysaccharide Utilization Loci. *Nucleic Acids Res* **46**, D677–D683 (2018).
42. Gotoh, A. *et al.* Sharing of human milk oligosaccharides degradants within bifidobacterial communities in faecal cultures supplemented with *Bifidobacterium*

bifidum. *Sci Rep* **8**, 13958 (2018).

43. Egan, M., Jiang, H., O'Connell Motherway, M., Oscarson, S. & van Sinderen, D. Glycosulfatase-encoding gene cluster in *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **82**, 6611–6623 (2016).

44. Bell, A. *et al.* Elucidation of a sialic acid metabolism pathway in mucus-foraging *Ruminococcus gnavus* unravels mechanisms of bacterial adaptation to the gut. *Nat Microbiol* **4**, 2393–2404 (2019).

45. Asakuma, S. *et al.* Physiology of consumption of human milk oligosaccharides by infant gut-associated bifidobacteria. *J Biol Chem* **286**, 34583–34592 (2011).

Methods

Chemicals

*p*NP-β-GlcNAc-6S and LNFP I were purchased from Carbosynth (Compton, UK). Sulfated Lewis[a] trisaccharide was purchased from Prozyme (Hayward, CA). *p*NP-β-GalNAc-6S and *p*NP-β-Gal-3S were from Tokyo Chemical Industry (Tokyo, Japan). *p*NP-β-GlcNAc-3S and *p*NP- β-GlcNAc-3,4-diS were gifts from Masanori Yamaguchi at Wakayama University (Japan). GlcNAc-6S, PGM, heparan sulfate (HS) from bovine kidney, and chondroitin sulfate A (CS) from bovine trachea were obtained from Sigma-Aldrich (MO, USA). Hyaluronic acid (HA) from *Streptococcus zooepidemicus* was obtained from Fuji-Film Wako Pure Chemicals (Osaka, Japan). Keratan sulfate (KS) from porcine shoulder cartilage was obtained from Iwai Chemicals Co. Ltd. (Tokyo, Japan). PCM was prepared as described previously[46]. Both PGM and PCM were dialyzed against water and lyophilized prior to use. All other chemicals used were of analytical grade.

Bacteria and culture conditions

B. bifidum JCM 1254 and *B. longum* JCM 31944 were obtained from the Japan Collection of Microorganisms (RIKEN BRC, Tsukuba, Japan) and cultured in Gifu Anaerobic Medium (Nissui Pharmaceutical, Tokyo, Japan) at 37 °C under anoxic conditions using an AnaeroPack (Mitsubishi Gas Chemical, Tokyo, Japan). *Escherichia coli* DH5α was used as a host for DNA manipulation. When necessary, antibiotics were added to the media as follows (μg/mL): ampicillin (Amp), 100; chloramphenicol (Cm), 20 for *E. coli* and 3 for *B. bifidum* and *B. longum*; spectinomycin (Sp), 75 for *E. coli* and 30 for *B. longum*.

Animal experiments

Conventional mice−C57BL/6N female mice at 9 weeks of age were purchased from Japan SLC (Shizuoka). The mice were housed individually in polycarbonate cages with bedding and given free access to drinking water and diet D12450H (Research Diets, NJ, USA) under controlled conditions of humidity (70%), lighting (12-h light/dark cycle), and temperature (22 ºC). *B. bifidum* cells [10^9 colony-forming units (CFU) in 200 μL phosphate-buffered saline (PBS)] or PBS only was administered by oral gavage to individual mice once a day for 5 consecutive days (n = 5/group). The experiment was commenced after a 3-day acclimation period. Body weight was measured as an indicator of food intake and health. Faecal pellets were collected from each mouse at day 0 and day 5 within 24 h post defecation, which were lyophilized and stored at −30 °C until use. At the end of experiment, animals were euthanized by cervical dislocation. Immediately after death, a midline incision was made to exteriorize the intestine and caecum. The intestines were flushed with PBS and scraped to collect mucus samples.

Germ-free and gnotobiotic mice−Germ-free ICR male mice, purchased from Sankyo Labo Service

(Tokyo, Japan), were randomly divided into three groups (n = 8) and housed in vinyl isolators under a 12-h light/dark cycle at 22 ºC. They were given free access to sterilized drinking water and diet D12450H. After 2-weeks acclimation period, the mice now at 7 weeks of age were administered 10^9 CFU of *B. bifidum* WT or *bbhII* mutant strain once by oral gavage to individual mice (n = 8/group) (day 0). At day 5, the mice were euthanized by anesthetic inhalation. Immediately after death, a midline incision was made to exteriorize the intestine and caecum. Faeces were collected from the colon (n = 7/WT group and n = 8/*bbhII* group) and stored at −30 °C until use. Sterile PBS was administered to the control group. One mouse in the control group died during the experiment (n = 7).

Inactivation of the *bbhII* gene in *B. bifidum*

The *bbhII* gene was inactivated by a single crossover recombination (Extended Data Fig. 5a). A suicide plasmid was constructed as follows. An In-Fusion HD Cloning kit (Takara Bio, Shiga, Japan) was used for ligation unless otherwise stated. First, the Sp^R gene of an *E. coli–Bifidobacterium* shuttle vector pMSK187, a derivative of pKKT427[47], was replaced with the Cm^R gene that was placed under the control of the *rpmB* promoter[48]. The Cm^R gene was synthesized at Thermo Fisher Scientific (Waltham, MA, US) (1260–1969 bp of pC194, GenBank accession: V01277.1), while the *rpmB* promoter was amplified by PCR from the *B. longum* genome. The resulting plasmid pMSK217 was used as a template for inverse PCR to remove the replicon for *Bifidobacterium*, which generated pMSK209. Finally, a PCR-amplified, 501 bp-internal region of *bbhII* was inserted into the NsiI site of pMSK209. The primers used are shown in Supplementary Table 9. The resulting plasmid, pHT33, which harbors pUC *ori*, the Cm^R gene, and a portion of the *bbhII* gene, was introduced into *B. bifidum* by electroporation[49] to allow for recombination at the *bbhII* locus. Inactivation of *bbhII* was confirmed by Western blotting using anti-BbhII antibodies and by measurement of activity using *p*NP-GlcNAc-6S as a substrate (Extended Data Fig. 5b,c).

Preparation of faecal extracts containing mucin glycoproteins

Extraction of mucin glycoproteins from faeces was carried out following the previously described method[7] with slight modifications. Briefly, freeze-dried faecal pellets (15−35 mg) were suspended in 20 volumes (*v*/*w*) of PBS and heated at 80 ºC for 15 min and then at 37 ºC for 90 min. The suspensions were centrifuged at 20,000 × *g* for 5 min and the supernatants (300 µL each) were transferred to new tubes. An equal volume of ice-cold ethanol was added and the tubes placed at −30 °C overnight. At this point the soluble fraction was pelleted by centrifugation at 20,000 × *g* for 15 min. Pellets were washed with 200 µL of ethanol and centrifuged again. The resulting precipitates were suspended in water and lyophilized. A portion of the resultant dried precipitates (200 µg) were subjected to glycomic analysis.

Release and purification of *O*-glycans from glycoproteins

The *O*-glycans were released by reductive β-elimination. Lyophilized samples (100 μg of PGM and PCM or 200 μg of faecal extracts) were reconstituted in 500 μL of 100 mM NaOH containing 1 M $NaBH_4$, and the mixtures incubated at 45 ºC for 18 h. The tubes were then placed on ice, to which aq 10% (*v*/*v*) acetic acid was added for neutralisation. An external standard consisting of 500 and 250 pmol of LNFP I and sulfated Lewis[a] trisaccharide, respectively, was added to the mixtures. The samples were then desalted with a Dowex-50W-X8 (H^+ form, 100−200 mesh, Sigma-Aldrich) column, followed by washing with 5% aq acetic acid (*v*/*v*). The flow-through and wash fractions were combined and lyophilized. Residual borate salts were removed as an azetrope by adding 0.3 mL of 10% acetic acid in methanol (*v*/*v*) and drying under a nitrogen gas stream at 40 ºC. This step was repeated five additional times. The released oligosaccharide alditols were dissolved in 0.3 mL of 5% aq acetic acid (*v*/*v*) and purified by a Sep-Pak C_{18} cartridge column (Waters, MA). The flow-through fraction was then lyophilized. The oligosaccharide alditols were reconstituted in 1 mL of water and further cleaned up using graphitized carbon columns (InertSep GC, GL Science, Tokyo, Japan). The samples (1 mL each) were applied onto the columns (150 mg/3 mL in size) that were pre-activated with 12 mL of aq 80% acetonitrile/0.1% trifluoroacetic acid (*v*/*v*) followed by equilibration with 8 mL of water. The loaded column was washed with 2 mL of water and eluted with 3 mL of aq 25% acetonitrile/0.05% trifluoroacetic acid (*v*/*v*). The eluates were lyophilized after removal of acetonitrile under a nitrogen gas stream at 40 ºC. The dried materials were subjected to glycan permethylation.

Permethylation and phase partition of *O*-glycans

Glycan permethylation and subsequent phase-partition of sulfated from non-sulfated *O*-glycans was performed as described previously[50] with some modifications. The lyophilized oligosaccharide samples were dissolved in 100 μL of anhydrous dimethyl sulfoxide (DMSO) and vigorously mixed with 250 μL of freshly prepared base (ca. 2.5 M NaOH in DMSO) and 150 μL of iodomethane for 5 min in glass tubes. After permethylation, 2 mL of 5% aq acetic acid (*v*/*v*) and 2 mL of dichloromethane were added. The tubes were vortexed and centrifuged at 470 × *g* for 3 min. Following transfer of the upper water-phase containing the sulfated *O*-glycans to a new tube, the *O*-glycans were separately purified using C18 SepPak cartridge column as mentioned above (Waters).

Mass spectrometric analysis of permethylated *O*-glycans

Molecular masses of permethylated glycans were analysed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) in the positive ion mode with an Autoflex III smartbeam (Bruker, Billerica, MA). 2,5-Dihydroxybenzoic acid was used as a matrix. MALDI-TOF/TOF MS was also performed to obtain MS/MS spectra of the precursor ion peaks. Only the precursors ion peaks that gave MS/MS fragments with predictable sugar compositions

were considered. Possible peeling products and underpermethylated ions were not considered.

Quantification of neutral monosaccharides and NeuAc

High-performance anion-exchange chromatography with pulsed amperometric detection was performed to quantify mucin *O*-glycan-constituting neutral monosaccharides and NeuAc in mouse caecum contents. A Dionex ICS-3000 (Thermo Fisher Scientific) system equipped with a CarboPac PA1 column (2 × 250 mm, Dionex) was used. For the separation of neutral monosaccharides, the elution was performed at a flow rate of 0.25 mL/min at 30 °C with an isocratic eluent of 14 mM NaOH and 5 mM CH_3COONa. For NeuAc detection, the elution was performed with a linear gradient of 0−330 mM CH_3COONa in 125 mM NaOH at 30 °C for 20 min. Standard curves were created using known concentrations of the respective carbohydrates.

GlcNAc-6S measurement

High-performance liquid chromatography (HPLC) analysis−GlcNAc-6S release from PGM was assessed by HPLC using fluorescence labelling with 2-aminoanthranilic acid (2AA) for detection[45]. To the reaction solutions, which contained 50 μM Gal as an internal control, equal volumes of 2AA-labeling solution [30 mg/mL 2AA, 20 mg/mL $NaBH_3CN$, and 4% $CH_3COONa(3H_2O)$ in methanol (*w*/*v*)] were added and the mixtures were incubated for 45 min at 80 °C. The GlcNAc-6S standard solutions were similarly labeled. The samples were then cooled to room temperature, mixed with five-fold volumes of acetonitrile, and loaded onto a Discovery DPA-6S SPE column (Sigma-Aldrich) pre-equilibrated with acetonitrile. After washing with 97% aq acetonitrile (*v*/*v*), the derivatized sugars were eluted in 1 mL of water. The eluates were analysed by a normal-phase HPLC (e2695, Waters) equipped with a TSKgel Amide-80 HR column (4.6 × 250 mm, Tosoh, Tokyo, Japan) at 65 °C. The column was equilibrated with 85% solvent A (acetonitrile)/15% solvent B (100 mM ammonium formate buffer, pH 4.3) and the elution was performed by a linear gradient of solvent B from 15% to 40% in 90 min at a flow rate of 1 mL/min. Fluorescence (Em. 420 nm; Ex. 330 nm) was monitored using a 2475 fluorescence detector (Waters). The peak areas of GlcNAc-6S were normalized by those of Gal for quantification.

Liquid chromatography (LC)-MS/MS analysis−Free GlcNAc-6S in mouse and human samples was measured using a LC-MS/MS system. Approximately 5 mg of caecum contents and freeze-dried faeces were homogenized in 100 μL of H_2O, to which 150 μL of phenol/chloroform/isoamyl alcohol [25:24:1 (*v*/*v*)] was added. Following vigorous shaking, the water phase was obtained by centrifugation and filtered through a 0.45 μm pore membrane (Millipore, MA, USA). A Prominence UFLC system (Shimadzu, Kyoto, Japan) with a Hypercarb column (2.1 × 100 mm, Thermo Fisher Scientific, MA, USA) kept at 45 ºC was used for separation. The mobile phase was a gradient between acetonitrile and 10 mM ammonium bicarbonate buffer (pH 10). The gradient consisted of of 5% acetonitrile (0−1 min),

5−30% (1−7 min), 30% (7−8 min), and 5% (8−11 min) at the flow rate of 0.2 mL/min. A triple quadrupole mass spectrometer (LCMS-8045; Shimadzu) equipped with a heated electrospray ionization probe was used for detection. The spectrometer was operated in the negative ion mode with the ion spray interface temperature at 300 ºC with argon gas used to obtain collision-induced dissociation. In the multiple reaction monitoring mode, the mass spectrometer detected ions by monitoring the decay of the *m/z* 300.10 precursor ion corresponding to the deprotonated molecule $[M-H]^-$ to the *m/z* 97.05 (collision energy [CE] = 28.0), 139.05 (CE = 26.0), 199.05 (CE = 17.0), and 282.10 (CE = 14.0) product ions each of which corresponds to $[OSO_3H]^-$, $[OCHCH_2OSO_3]^-$, ^{0,2}A ring-cleavage product, and $[M-H_2O-H]^-$, respectively. 6-*O*-Sulfated *N*-acetylhexosamine was shown to provide a characteristic dehydrated ion $[M-H_2O-H]^-$ at *m/z* 282.10, which enabled discrimination between 6-*O*-sulfated and 3/4-*O*-sulfated *N*-acetylhexosamines[51]. GlcNAc-6S and GalNAc-6S were separated by LC with different retention times (GlcNAc-6S for 2.8 min; GalNAc-6S for 3.3 min). The standard curve of GlcNAc-6S was linear in the range between 0.0195 μM and 1.25 μM for all the fragment ions as well as in the total ion chromatogram.

GAG degradation ability of *B. bifidum*

B. bifidum cells grown overnight were harvested by centrifugation and suspended in PBS to give an OD_{600} value of 0.8, to which the same volume of 0.8% GAGs (HS, KS, CS, or HA) or PGM (*w*/*v*) was added. The mixtures were incubated at 37 ºC for 24 h, and aliquots were used for thin-layer chromatographic analysis (Silica Gel 60, Sigma-Aldrich) with a solvent system of *n*-butanol/acetic acid/water [2:1:1 (*v*/*v*)]. The sugars were visualized using a diphenylamine-aniline-phosphoric acid reagent[42].

Human stool sample collection

Stool samples were obtained from healthy Japanese infants (15 male and 18 female, age range of 0.2−13 months old) and adults (8 male and 10 female, age range of 10s−60s years old). Infant samples were collected at Nagao Midwife Clinic (Kyoto, Japan). The samples were frozen at the clinic and transferred to the laboratory. Adult samples were self-collected and immediately transferred to the laboratory. Small portions (~ 1 g) of five adult fresh samples out of the total of 18 were separately taken, washed with anoxic PBS, suspended in 20% glycerol in an anoxic chamber Invivo$_2$ 400 (Ruskinn Technology, Bridgend, UK; 10% CO_2, 10% H_2, and 80% N_2), and stored at −80 ºC for faecal cultivation. All the other samples were lyophilized and subjected to DNA extraction and LC-MS/MS analysis.

Faecal DNA extraction

Approximately 50 mg of freeze-dried faeces was suspended in 300 μL of InhibitEX buffer appended

to a QIAmpFast DNA Stool Mini Kit (Qiagen, Hilden, Germany) and the mixtures were incubated at 95 °C for 10 min. The suspension was transferred to a new tube containing one 5.0 mm-stainless bead and approximately 200 mg of 0.1 mm-zirconia beads, and the tube was vigorously shaken at 1,500 rpm for 10 min with a Shake Master NEO (Bio Medical Science, Tokyo, Japan). The suspensions were again heated at 95 °C for 10 min, centrifuged, and the resultant supernatants (200 μL) were subjected to conventional phenol-chloroform extraction [phenol/chloroform/isoamyl alcohol = 25:24:1 (*v*/*v*)], which was followed by ethanol precipitation to obtain faecal DNA.

Microbiota analysis

16S rRNA gene-based microbiota analysis was conducted as follows. The V3−V4 region was amplified using a two-step PCR approach with Takara Ex Taq Hot Start Version polymerase (Takara Bio). The first PCR, which contained 0.2 μM of each primer and 1 μL of faecal DNA (20~60 ng/μL) in a total volume of 20 μL, consisted of 30 cycles of 94 °C for 30 s, 50 °C for 30 s, and 72 °C for 30 s. The reactions were performed in three separate tubes for each sample. The products from the first PCR tubes were combined and an aliquot (1 μL) of the mixture was used as a template for the second PCR. The second PCR was run with 8 cycles under the same conditions as the first, but using a different primer pair. The primers used are indicated in Supplementary Table 9. The sequencing was performed using an Illumina MiSeq instrument with a MiSeq v3 Reagent kit (Illumina, CA, US). Sequences consistent with data from the Genome Reference Consortium human build 38 (GRCh38) and phiX reads were removed from the raw Illumina paired-end reads. The sequences were then analysed using the QIIME2 software package, version 2017.10 (https://qiime2.org/). Potential chimeric sequences were removed using DADA2[52] after trimming 30 and 90 bases of the 3′-region of the forward and reverse reads, respectively. Taxonomic classification was performed using a Naive Bayes classifier trained on Greengenes 13.8 database clustered at 99% identity threshold in the entire 16S rRNA gene. Weighted UniFrac distances were calculated using QIIME2. Comparison of bacterial taxon abundance were performed by the Linear discriminant analysis Effect Size (LEfSe)[53] algorithm with the default settings.

Quantitative PCR (qPCR)

qPCR was performed with a Thermal Cycler Dice Real-Time System (TaKaRa Bio). Each reaction mixture (total volume of 15 μL) contained 7.5 μL 2 × TB Green Premix EX Taq™ II (Tli RNaseH plus) (TaKaRa Bio), 0.6 μL (10 pmol each) of primer pairs (Supplementary Table 9), and 6.9 μL of the extracted faecal DNA solution. The reaction consisted of an initial denaturation of 30 s at 95 °C followed by 45 cycles of 95 °C for 5 s, 55 °C for 10 s, and 72 °C for 20 s. Melting curves were generated after the cycles to verify the specific amplifications. Known concentrations of the genomic DNA of *B. bifidum* were used for creating calibration curves of *bbhII* and total 16S rRNA genes. The

lowest detection limits of *bbhII* and 16S rRNA gene were 5.4×10^3 and 6.1×10^4 copies/μg DNA, respectively.

Human stool cultivation

Thawed stool suspensions (20% glycerol) were washed three times with basal medium containing 4% reducing solution[15] and suspended with the same medium supplemented with and without 10 mM GlcNAc-6S. The suspension was incubated at 37 °C for 24 h under anoxic conditions. DNA was extracted pre and post cultivation and used for microbiota analysis.

Metagenomic dataset mining

A publicly available metagenomic dataset[21] was used for the analysis. The data of the 4-month-old infants who were not administered any antibiotics and were breastfed or mixed-fed until 4 months old were extracted. The dataset obtained from their mothers were also extracted. Consequently, the data of 80 infants and 80 mothers were used for the analysis. The sequence data were downloaded from the ENA server (https://www.ebi.ac.uk/ena/browser/home). Species identification and quantification was done using the standard Kraken2 database (version 2.1.2)[54], which was generated through the default kraken2-build command. The standard database was chosen over the bacterial database to allow for the identification of human contamination. As a result, 41.1 ± 9.28% (mothers) and 74.1 ± 12.2% (infants) of total reads were annotated to specific taxa. For gene quantification, the BLAST+ application (version 2.10.1)[55] was used. A protein-based BLAST search was performed to quantify *sgl* and *bbhII* homologues using the tBLASTn function of the BLAST+ application. A stringent high identity search (> 90% in both identity and coverage) was performed including the protein sequences of all presumed homologs. The Sgl homologue clade includes Sgl and BT_4394, while the BbhII homologue clade contains BbhII, WP_172192827.1, WP_153878949.1, CRH87835.1, WP_206666329.1, WP_125968884.1, and WP_076060111.1 (Extended Data Fig. 3). Spearman's rank correlation analysis between the relative abundances of each bacterial taxon and the Sgl or BbhII homologue was performed using R ver. 4.0.5, followed by Benjamini–Hochberg false discovery rate (FDR) correction[56] using GraphPad Prism 8.4.3.

Recombinant BbhII expression in *E. coli*

Plasmid pET23b(+)–*bbhII*, which harbors the gene encoding 39−1027 aa of BbhII with a C-terminal His_6-tag, was constructed previously[15] and used as the PCR template for generating BbhII variants (Supplementary Fig. 3). QuikChange methodology was employed for introducing amino acid replacements. Deletion mutants were created by normal or inverse PCR followed by ligation using an In-Fusion HD Cloning kit (Takara Bio). Primers used are listed in Supplementary Table 9. All constructs generated by PCR-based techniques were sequenced to ensure that no base change other

than those planned had occurred.

E. coli BL21(DE3) Δ*lacZ*-CodonPlus cells containing *bbhII* variants on plasmid were cultivated in LB medium containing Amp and Cm at 18 °C. When the OD_{600} reached 0.5, 0.1 mM isopropyl-β-D-thiogalactopyranoside was added. Following further incubation for 24 h at 18 °C, the cells were harvested, suspended in lysis buffer [50 mM HEPES {4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid} (pH 8.0), 300 mM NaCl, and 10 mM imidazole], and disrupted by sonication. After centrifugation, the resulting supernatant was applied to a Ni-NTA spin column (Qiagen). Purification was done according to the manufacturer's protocol. The eluate was collected and desalted using an Amicon Ultracel-10K centrifugal filter (Millipore) and applied to a Mono Q 5/50 GL column (GE Healthcare, Little Chalfont, UK) preequilibrated with 20 mM Tris-HCl (pH 8). The elution was performed by a linear gradient of 0−0.5 M NaCl in the same buffer. The proteins were further purified using a Superdex 200 Increase 10/300 GL column (GE Healthcare) pre-equilibrated with 10 mM Tris-HCl (pH 8.0) containing 300 mM NaCl. The pure fractions were combined and concentrated as above. The purity was evaluated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) followed by Coomassie Brilliant Blue staining (Quick-CBB, Wako Pure Chemical). The protein was quantified using a theoretical absorption coefficient at 280 nm, calculated based on the sequence (https://web.expasy.org/protparam/).

Inhibitor synthesis

1,2-dideoxy-2'-methyl-α-D-glucopyranoso-[2,1-d]-Δ2'-thiazoline-6-sulfate (NAGT-6S) sodium salt

This compound was prepared according to the procedure of Liu *et al.*[57]

O-*(2-Acetamido-2-deoxy-6-*O-*sulfo-*D-*glucopyranosylidene)amino* N-*phenylcarbamate (PUGNAc-6S) sodium salt*

Sulfur trioxide trimethyl amine complex (22 mg, 0.16 mmol) was added to PUGNAc[29,59] (51 mg, 0.14 mmol) in pyridine (2 mL) at 0 °C and the mixture stirred (0 °C, 2 h) then left overnight at room temperature. The reaction mixture was quenched at 0 °C with aq. 1 M $NaHCO_3$ (0.35 mL, 0.35 mmol) solution and concentrated. Purification of the resultant residue by flash column chromatography (MeOH:CH_2Cl_2 3:7) gave the desired compound as a white solid (25 mg, 39%). Additionally 31 mg of starting material was recovered. R_f = 0.33 (MeOH:CH_2Cl_2 3:7). ^{1}H NMR (500 MHz, CD_3OD): δ 7.50-7.44 (m, 2H), 7.32-7.26 (m, 2H), 7.07-7.02 (m, 1H), 4.64-4.59 (m, 1H), 4.42 (dd, *J* = 2.1, 11.6 Hz, 1H), 4.32 (dd, *J* = 4.5, 11.6 Hz, 1H), 4.23-4.17 (m, 1H), 3.82-3.75 (m, 2H), 2.06 (s, 3H); ^{13}C NMR (125.8 Hz, CD_3OD): δ 173.8, 158.8, 154.7, 139.4, 129.9, 124.6, 120.2, 81.5, 74.3, 69.9, 67.2, 52.9, 22.8; FTIR (ATR): ν = 3290 (m), 1749 (m), 1644 (m) cm^{-1}; HR-MS (ESI-): m/z [M]$^-$ calcd. for $C_{15}H_{18}N_3O_{10}S$: 432.0713, found: 432.0720. See Supplementary Figure 5.

Crystallography

Purified BbhII WTc-His_6 was used for crystallization. Selenomethionine (SeMet)-labelled protein was prepared by the method as described previously[58] and purified similarly as described for the native protein. Crystals were grown at 20 ºC using the sitting-drop vapor diffusion method, by mixing 0.5 μL of the protein solution containing 11 mg/mL BbhII and 10 mM ligand (GlcNAc-6S or PUGNAc-6S) with an equal volume of reservoir solution consisting of PEG 8000 (*w*/*v*) and 0.1 M HEPES-NaOH (pH 7.5) for the GlcNAc-6S complex or 20% (*w*/*v*) PEG 3000 and 0.1 M sodium citrate (pH 5.5) for the PUGNAc-6S complex. Crystals of SeMet-BbhII complexed with GlcNAc-6S were grown similarly using a reservoir solution containing 20% (*w*/*v*) PEG 3350 and 0.2 M KCl. For cryoprotection, 20% 2-methyl-2,4-pentanediol and 20% trehalose were used for the GlcNAc-6S and the PUGNAc-6S complexes, respectively. The crystals were flash-cooled by dipping into liquid nitrogen. Diffraction data were collected at 100 K on beamlines at SPring-8 (Hyogo, Japan) and the Photon Factory of the High Energy Accelerator Research Organization (KEK, Tsukuba, Japan). The data collection and refinement statics for crystallography are shown in Supplementary Table 5 with the software used. Molecular graphic images were prepared using PyMOL (Schrödinger, LLC, New York, NY, USA).

Enzyme assay

Sulfoglycosidase activity was routinely assessed using *p*NP-β-GlcNAc-6S. The reaction was carried out at 37 °C in 50 mM sodium citrate buffer (pH 5.5) containing 2 mM substrate in the presence of purified BbhII variants (2.5 μg/mL). Reactions were stopped by adding 5 volumes of 1 M Na_2CO_3 and the absorbance at 405 nm was measured to quantify *p*-nitrophenolate. Assays were performed in which the linearity of the reaction rate was observed. Inhibition by PUGNAc-6S and NAGT-6S was examined in assays consisting of 100 mM sodium citrate (pH 5.5), 0.2−1.0 mM *p*NP-β-GlcNAc-6S, the inhibitor (0−100 nM PUGNAc-6S or 0−250 nM NAGT-6S), and 2 μg/mL BbhII WTc-His_6. The reaction mixture was aliquoted and stopped every 5 min by adding 4 volumes of 0.1 M NaOH. The kinetic parameters were calculated by curve-fitting the experimental data to the competitive inhibition equation using SigmaPlot 12.0 (HULINKS, Tokyo, Japan).

When PCM or PGM was used as a substrate, the glycoprotein (800 μg) was suspended in 50 mM sodium citrate buffer (pH 5.5), and the mixture was incubated at 37°C for 16 h in the presence of either the purified BbhII variants (1 μM) or the recombinant *B. longum* cells expressing BbhII variants in a total volume of 100 μL. The reaction was ceased by adding ice-cold acetone (400 μL) and the mixture was placed on ice for 15 min. After centrifugation, the precipitated protein was suspended in 800 μL of water and aliquots (100 μL) were subjected to *O*-glycan analysis.

NAGT-6S-mediated inhibition of mucin *O*-glycan breakdown by *B. bifidum*

Pre-cultured *B. bifidum* cells were harvested by centrifugation, washed, and resuspended in 4% reducing solution[15] to give an OD_{600} value of 0.5. The suspension was then used to inoculate the basal

media[15] supplemented with 0.4% PGM to give an OD_{600} of 0.05. The cultures were incubated in the absence and presence of 0.1 or 1 mM NAGT-6S under the anoxic conditions at 37 °C for 24 h and subjected to *O*-glycan analysis.

Isothermal titration calorimetry (ITC)

Binding thermodynamics were analysed using an MicroCal ITC200 isothermal titration calorimeter (Malvern Panalytical, Malvern, UK) at 30 ºC ± 0.1 ºC. CBM-His_6 and its W183A variant were used for the analysis (Supplementary Fig. 3a). The proteins were dialyzed against PBS, and the dialysis buffer was used to dissolve the ligands. Protein concentrations were adjusted to 100 μM. The proteins were titrated with ligand (1 mM) by 20 injections (0.2 μL first followed by 2 μl) with the dilution heat being negligible. The data were analysed using Origin 7.0 software by fitting to one set of sites model.

Preparation of anti-BbhII antibodies and Western blotting

Rabbit antiserum against BbhII was prepared by Eurofin Genomics (Tokyo, Japan). Purified BbhII WT-His_6 [1.0 mg/mL in 20 mM Tris-HCl (pH7.4) and 15 mM NaCl] was used for immunization. The antibodies were purified from the serum by using agarose-beads (AminoLink plus coupling resin, Thermo Fischer Scientific) that were conjugated with BbhII. The conjugation and purification were carried out according to the manufacturer's instruction. Rabbit anti-BbhII antibodies and anti-GLBP (galacto-*N*-biose/lacto-*N*-biose I-binding protein) antibodies[45] were used as the primary antibodies with 10,000-fold and 20,000-fold dilutions, while goat anti-rabbit IgG-HRP conjugate (Santa Cruz Biotechnology, TX, USA) was used as the secondary antibody, at a 10,000-fold-dilution.

ELISA

Indirect ELISA was performed to evaluate the binding of BbhII CBM-His_6 to mucin *O*-glycans. PGM and PCM were serially diluted with PBS to 0.3125−20 μg/mL, and 100 μL of the resultant samples were applied to a 96-well Maxisorp Nunc-Immuno ELISA Clear Plate (Thermo Fischer Scientific). The plate was kept for 4 h at room temperature to immobilize the mucins. The wells were subsequently washed three times with 600 μL of TBS-T [100 mM Tris-HCl (pH7.4) containing 150 mM NaCl and 0.1% Tween-20] and blocked by 100 μL of Blocking One (Nacalai tesque, Kyoto, Japan) at 4 ºC overnight. The wells were washed again with TBS-T three times, and to which 100 μL of CBM-His_6-complex agent [15.5 nM CBM-His_6, 0.33 μg/mL Penta-His mouse IgG (Qiagen), and 0.02 μg/mL secondary goat anti-mouse IgG-HRP (Santa Cruz Biotechnology) pre-incubated for 30 min in 5 mL Blocking One] was added. After incubation at 4 ºC overnight, the wells were washed with TBS-T three times. The binding of CBM-His_6 to mucin *O*-glycans was detected by adding 100 μL of 1-Step™ Turbo TMB-ELISA substrate solution (Thermo Fischer Scientific). One hundred μL of 1 M HCl was used to stop the reaction, and the absorbance at 450 nm was measured. Sugars were added at the final

concentration of 0.25 mM when examining binding inhibition.

Heterologous expression of BbhII variants in *B. longum*

The DNA region containing the *bbhII* open reading frame and its upstream (200 bp) and downstream (100 bp) flanking regions was PCR-amplified from the *B. bifidum* genome and inserted into NdeI site of *E. coli-Bifidobacterium* shuttle vector pTK2064[59], to generate pTK2064-*bbhII* (WT). The W183A and ΔCBM variants were created by QuikChange methodology and inverse-PCR using pTK2064-*bbhII* (WT) as a template, respectively. The primers used are listed in Supplementary Table 9. After sequence confirmation, the resulting plasmids were introduced into *B. longum* by electroporation[49]. The transformants were selected for on Cm-containing agar plates.

CAZy database search

A CAZy database was used to analyse the prevalence and abundance of GHs and CBMs in the genomes of *B. bifidum*, *C. perfringens*, *A. muciniphila*, *B. caccae*, *B. fragilis*, *B. thetaiotaomicron*, *P. melaninogenica*, and *R. gnavus*. The GHs and CBMs presumed to be associated with mucin *O*-glycan degradation and interaction, i.e. GH2, 16, 20, 29, 31, 33, 35, 35, 36, 42, 84, 89, 95, 98, 101, 109, 110, 112, 123, 129, and 136, and CBM32, 40, 47, 51, and 71, were then extracted from the database (Supplementary Table 8). To determine the effect size and significance of the presence or absence of *muc*-GHs on the distribution of *muc*-CBMs within the 199 strains belonging to the above 8 species, a PERMANOVA with 9,999 iterations was performed using the 'envfit' function in the package 'vegan' (community ecology package). NMDS on the distribution of *muc*-CBMs was used for ordination based on Bray-Curtis distances. Statistical analysis was performed using R ver. 4.1.1.

Ethical consideration

Animal experiments were approved by the Kyoto University Animal Experimentation Committee (Lif-K20021 and Lif-K21020) and performed in June of 2020 and July of 2022. The experiments using human samples were reviewed and approved by the Ethics Committees of Kyoto University (R0046) and the University of Shiga Prefecture (71-3) and were performed according to the Declaration of Helsinki. Written informed consent was obtained from all individuals except those under 18 years old, for whom their mother's consent was obtained.

Statistics

Statistical analyses were performed with R ver 4.0.5 or 4.1.1 and GraphPad Prism 9.4.1. *P* values of less than 0.05 were considered statistically significant.

Data Availability Statement

Atomic coordinates and structure factors of GlcNAc-6S- and PUGNAc-6S-complexed BbhII proteins from *B. bifidum* JCM 1254 have been deposited in the PDB under accession numbers 7WDT and 7WDU, respectively. The sequences of 16S rRNA V3-V4 variable regions of faecal microbiotas of mice and humans have been deposited in the DDBJ databank under the accession numbers DRA013515 and DRA013516, respectively.

Material Availability Statement

All biological materials, except for human and mouse samples, are publicly available or will be distributed upon request.

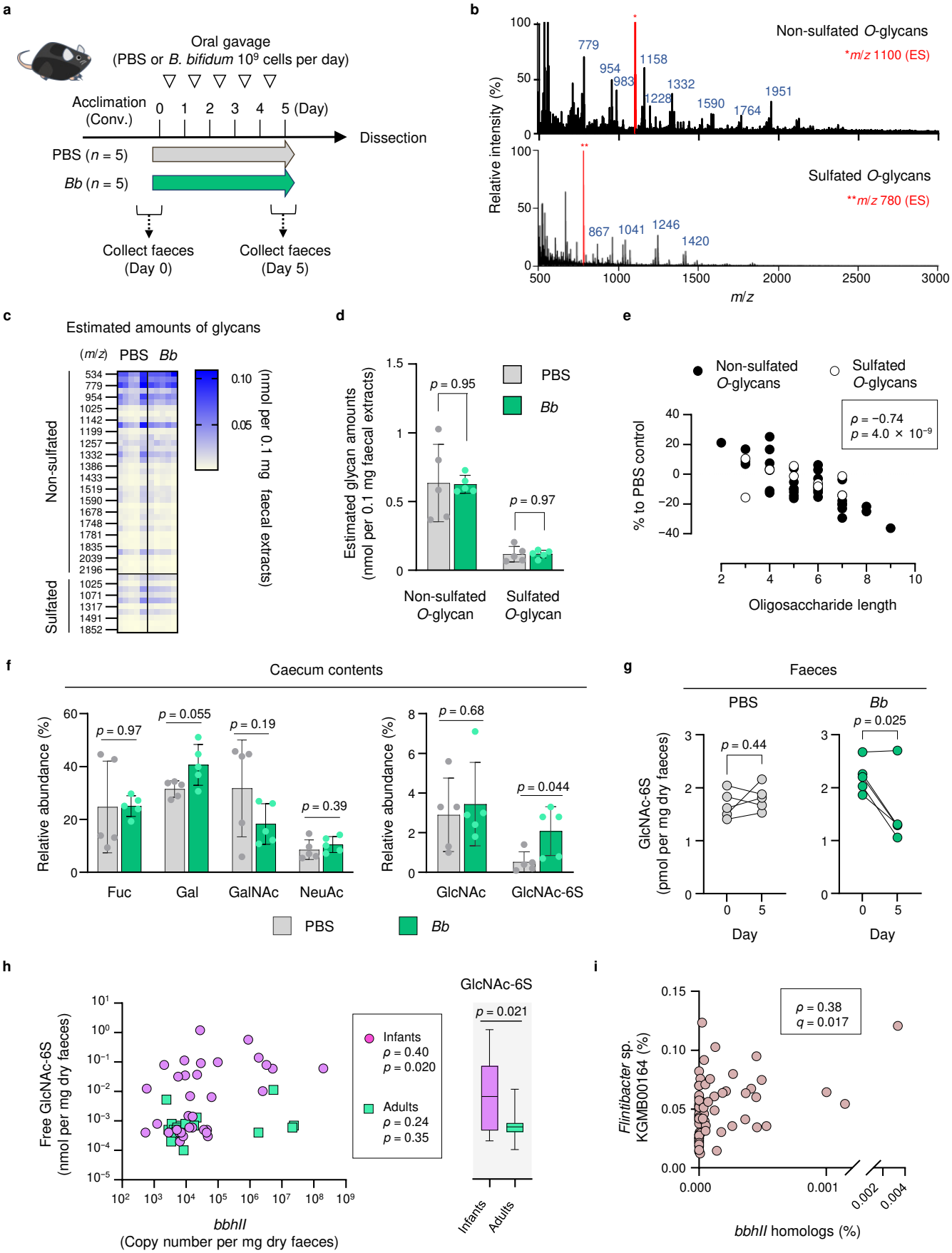
Code Availability Statement

No custom code was used in this study.

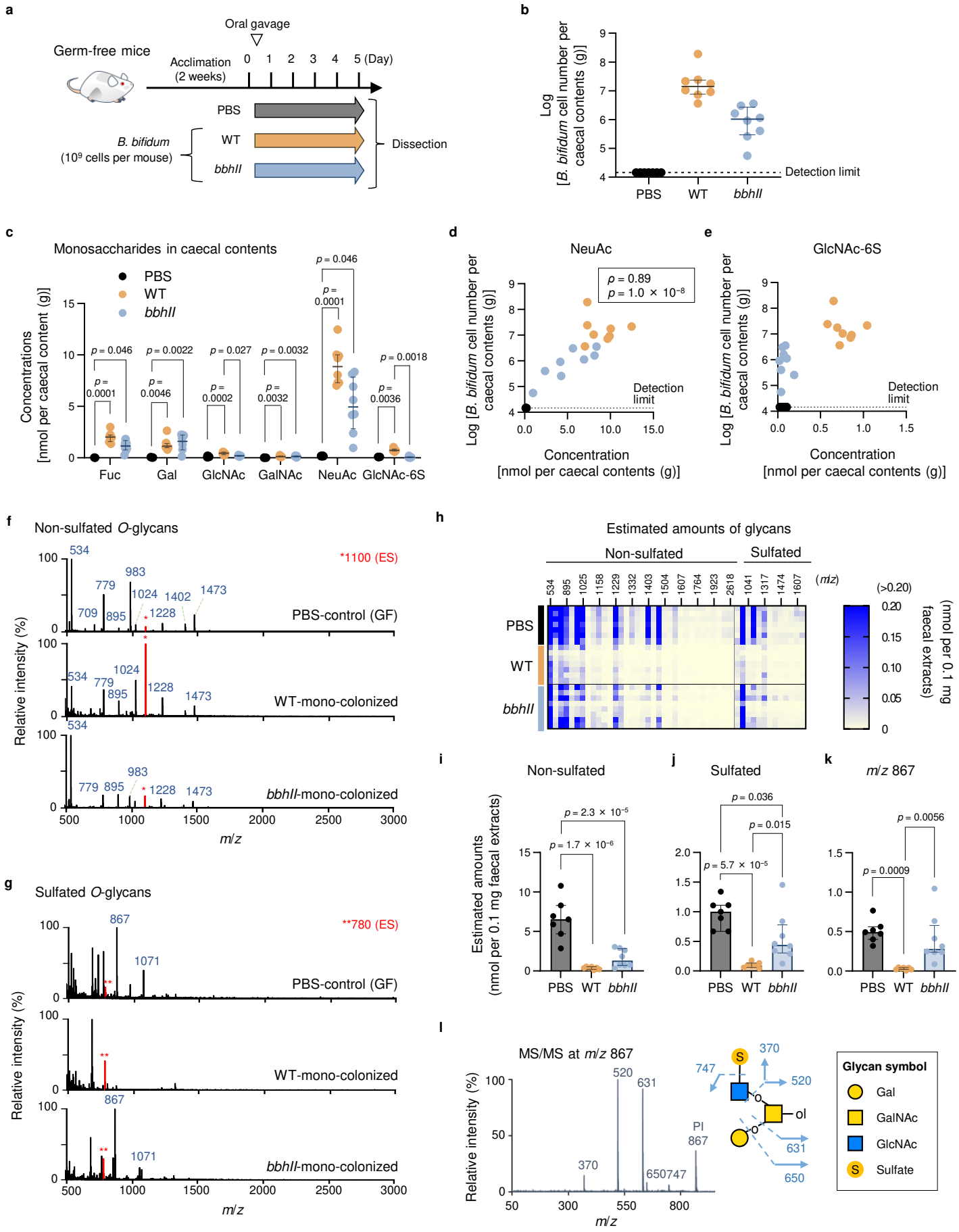
Methods-only references

46. Nishiyama, K. *et al.* Evaluation of bifidobacterial adhesion to acidic sugar chains of porcine colonic mucins. *Biosci Biotechnol Biochem* **78**, 1444–1451 (2014).
47. Yasui, K. *et al.* Improvement of bacterial transformation efficiency using plasmid artificial modification. *Nucleic Acids Res* **37**, e3–e3 (2009).
48. Kozakai, T., Shimofusa, Y., Nomura, I. & Suzuki, T. Construction of a reporter system for bifidobacteria using chloramphenicol acetyltransferase and its application for evaluation of promoters and terminators. *Biosci Microbiota Food Health* **40**, 115–122 (2021).
49. Sakanaka, M. *et al.* Evolutionary adaptation in fucosyllactose uptake systems supports bifidobacteria-infant symbiosis. *Sci Adv* **5**, eaaw7696 (2019).
50. Kumagai, T., Katoh, T., Nix, D. B., Tiemeyer, M. & Aoki, K. In-gel β-elimination and aqueous-organic partition for improved *O*- and sulfoglycomics. *Anal Chem* **85**, 8692–9 (2013).
51. Minamisawa, T. & Hirabayashi, J. Fragmentations of isomeric sulfated monosaccharides using electrospray ion trap mass spectrometry. *Rapid Commun Mass Spectrom* **19**, 1788–1796 (2005).
52. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**, 581–583 (2016).
53. Segata, N. *et al.* Metagenomic biomarker discovery and explanation. *Genome Biol* **12**, R60 (2011).
54. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol* **20**, 257 (2019).
55. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

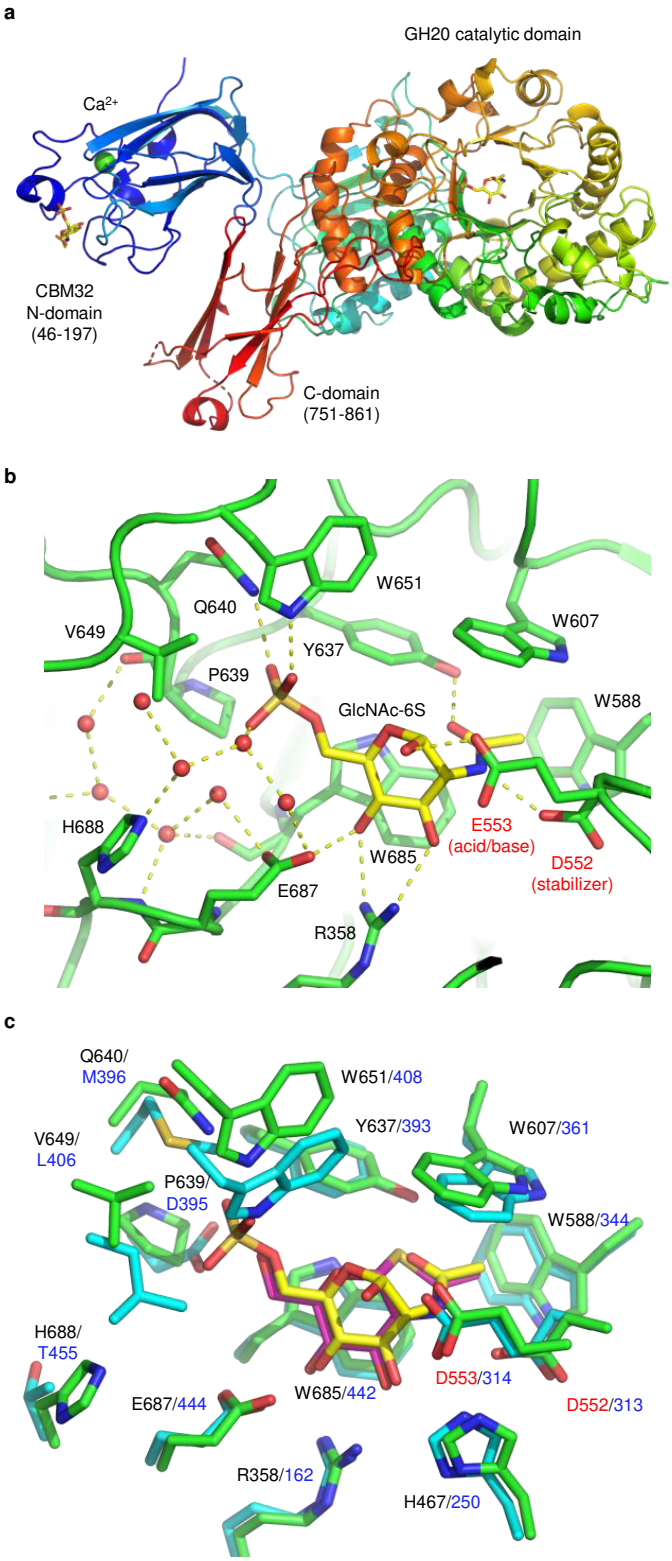
56. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289–300 (1995).
57. Liu, J., Shikhman, A. R., Lotz, M. K. & Wong, C. H. Hexosaminidase inhibitors as new drug candidates for the therapy of osteoarthritis. *Chem Biol* **8**, 701–711 (2001).
58. 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).
59. 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).



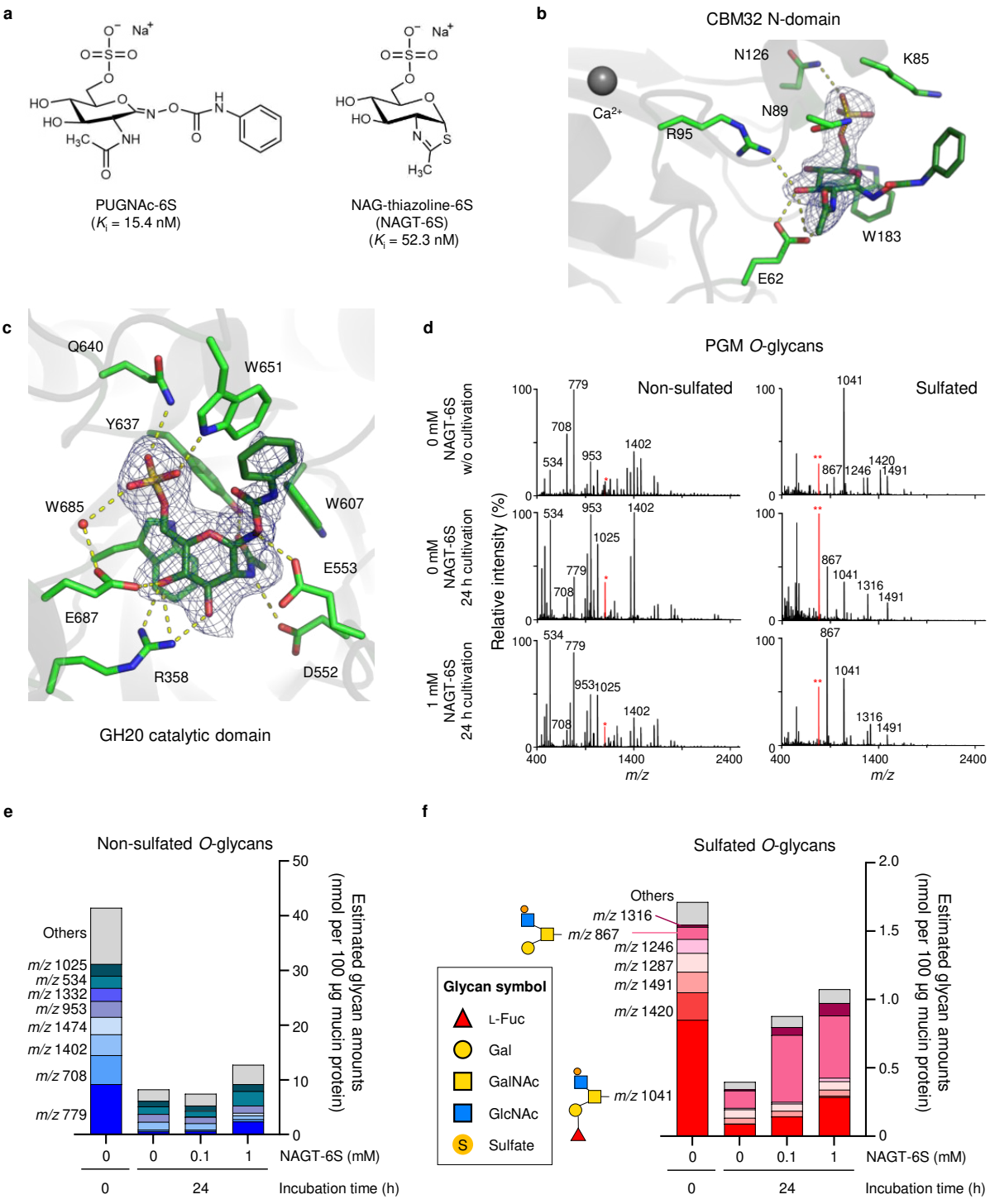
Katoh *et al.*, Figure 1



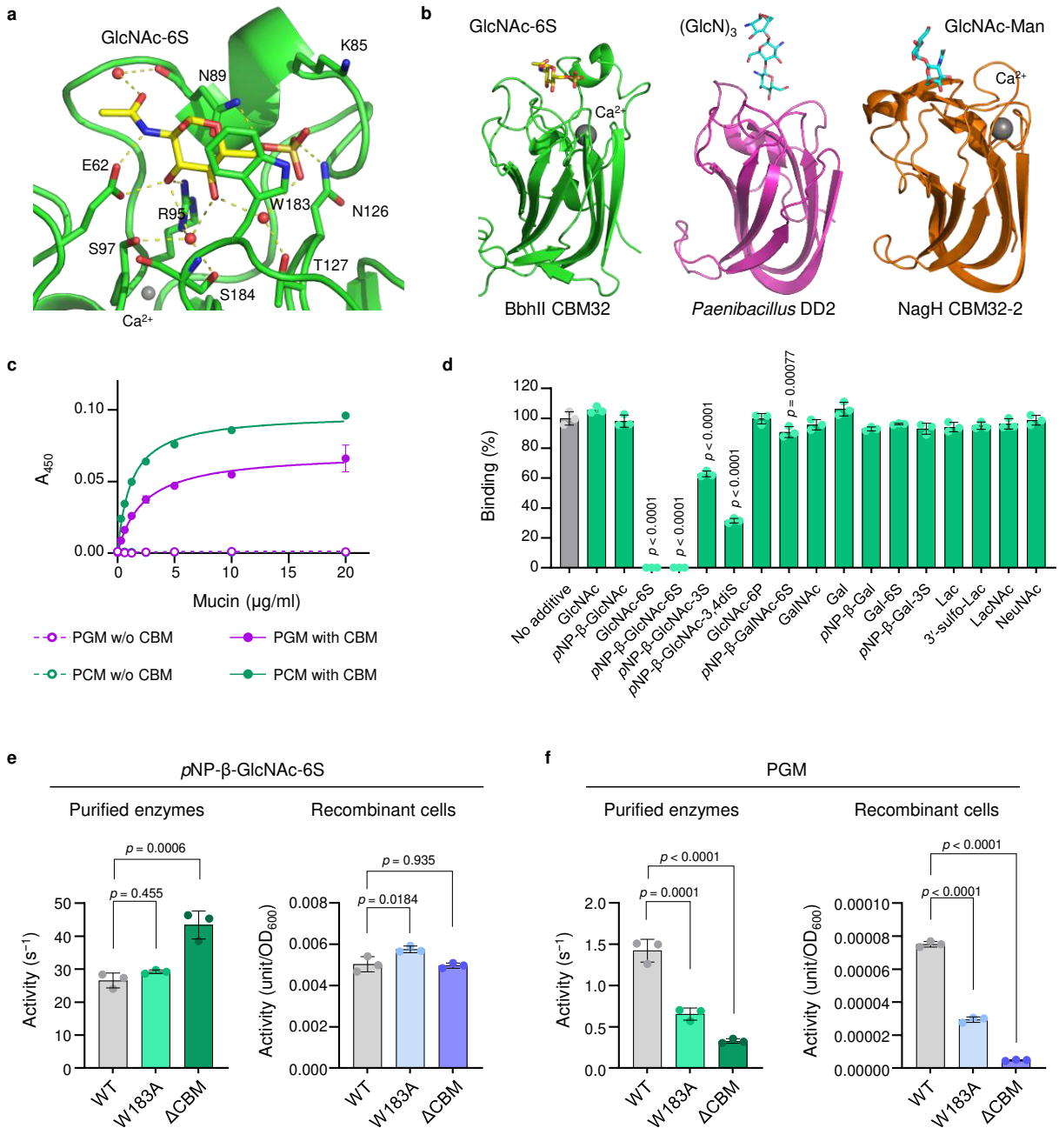
Katoh *et al.*, Figure 2



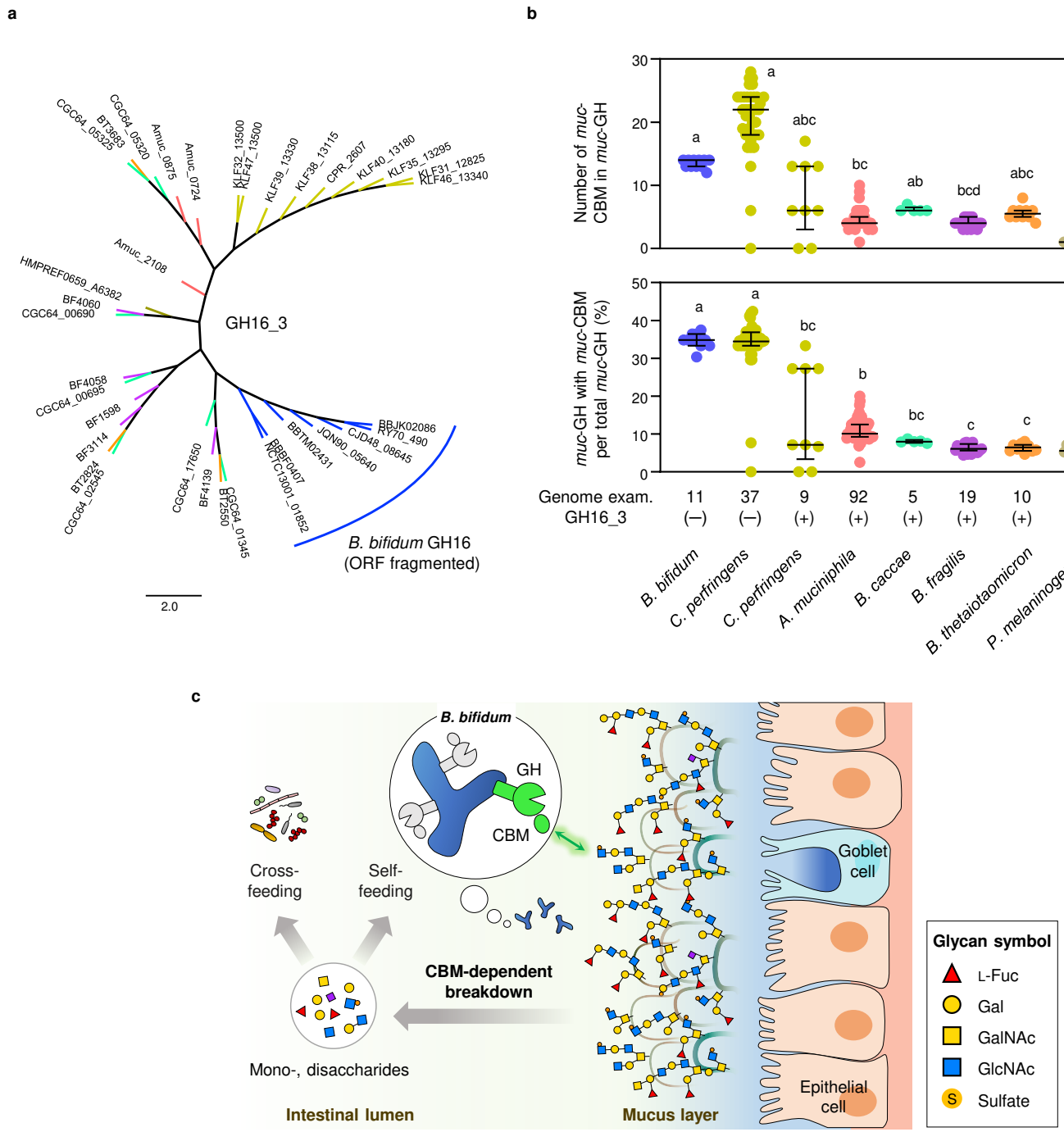
Katoh *et al.*, Figure 3



Katoh *et al.*, Figure 4

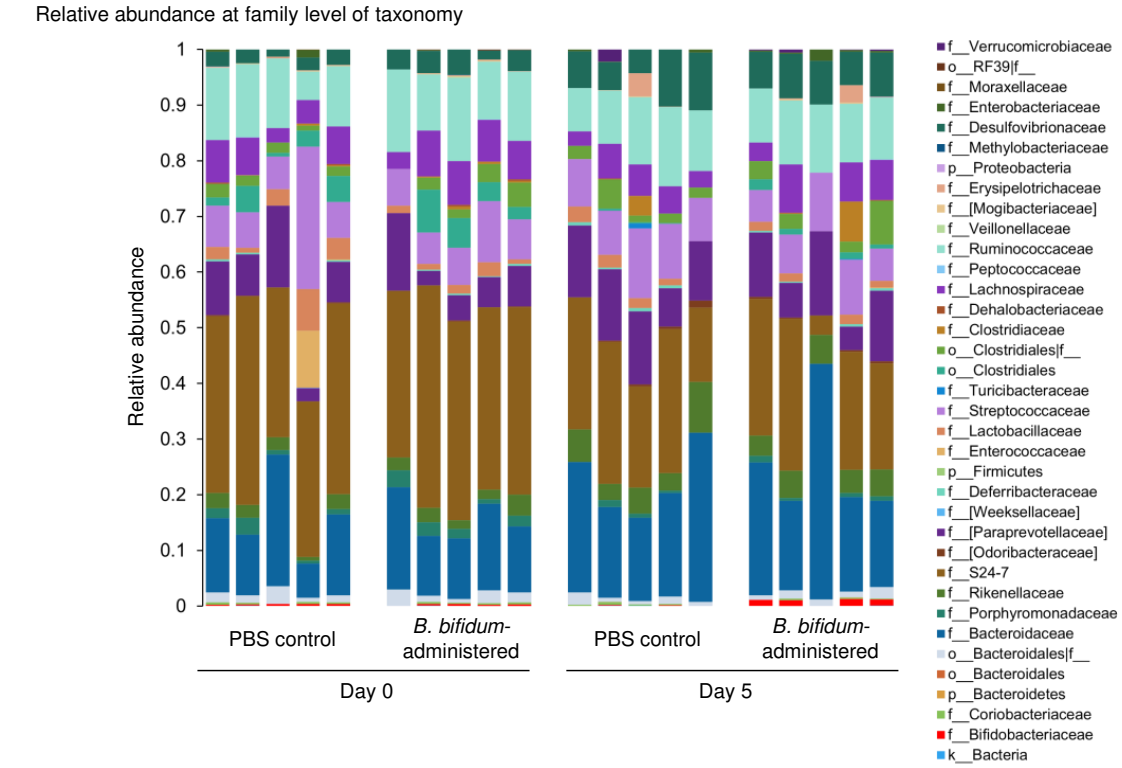


Katoh *et al.*, Figure 5

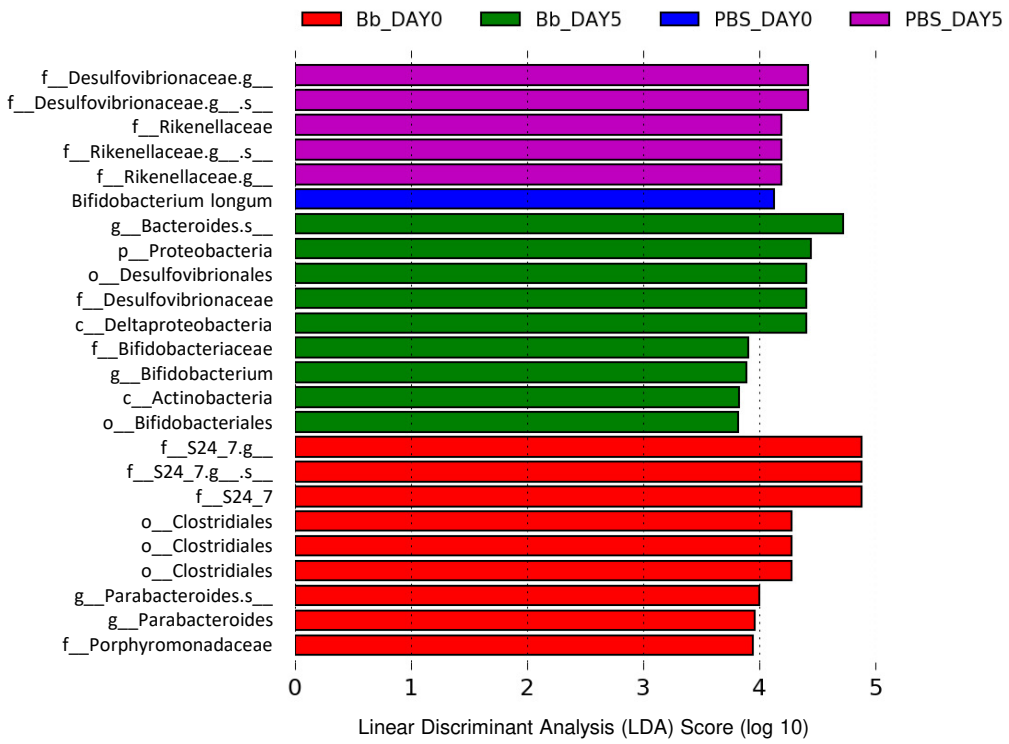


Katoh *et al.*, Figure 6

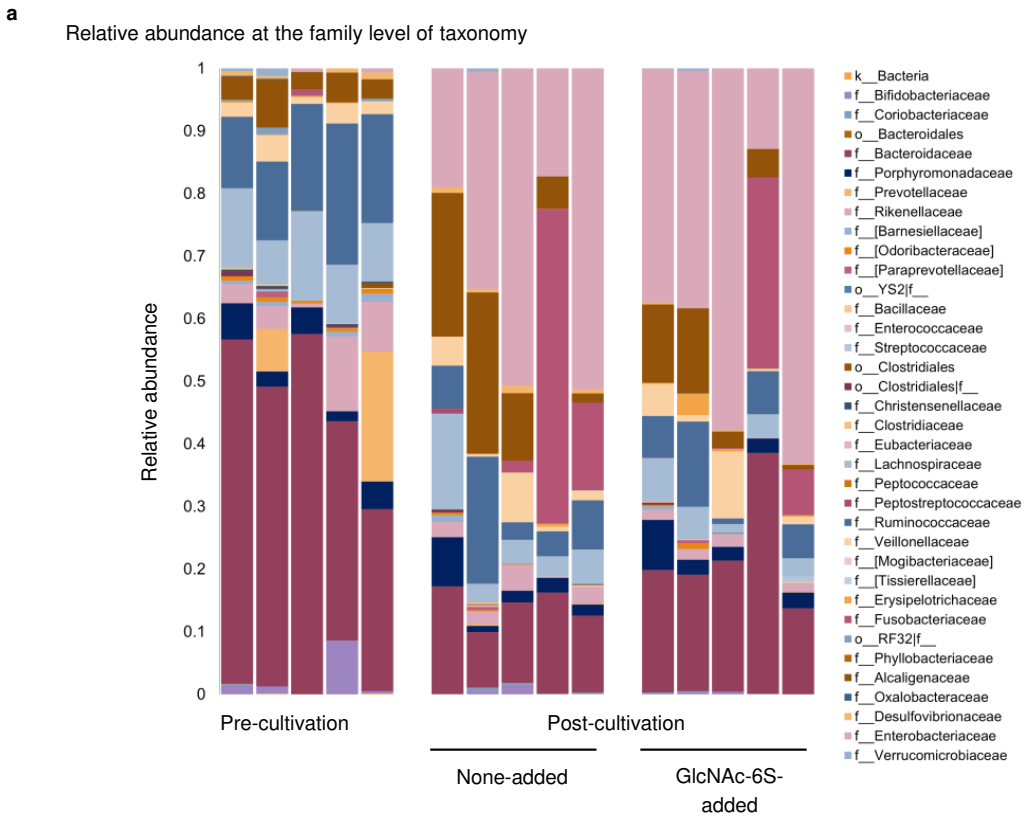
a

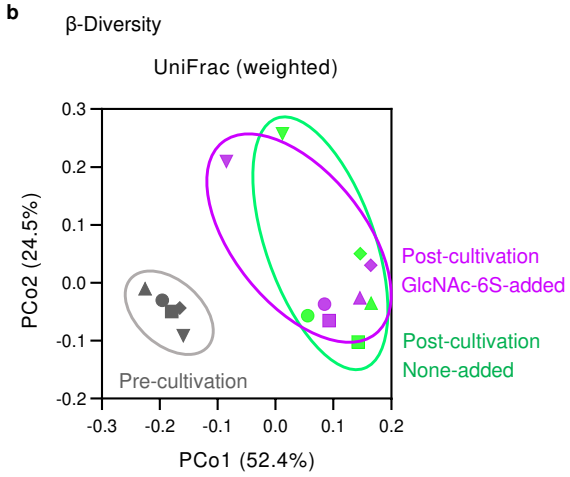


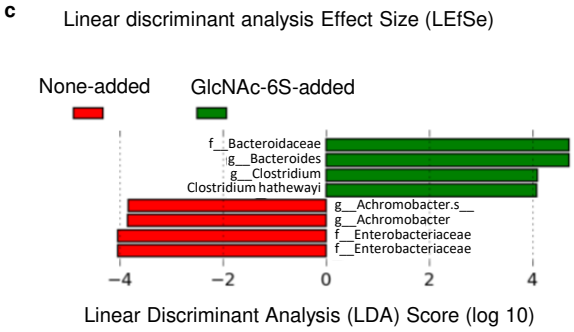
b



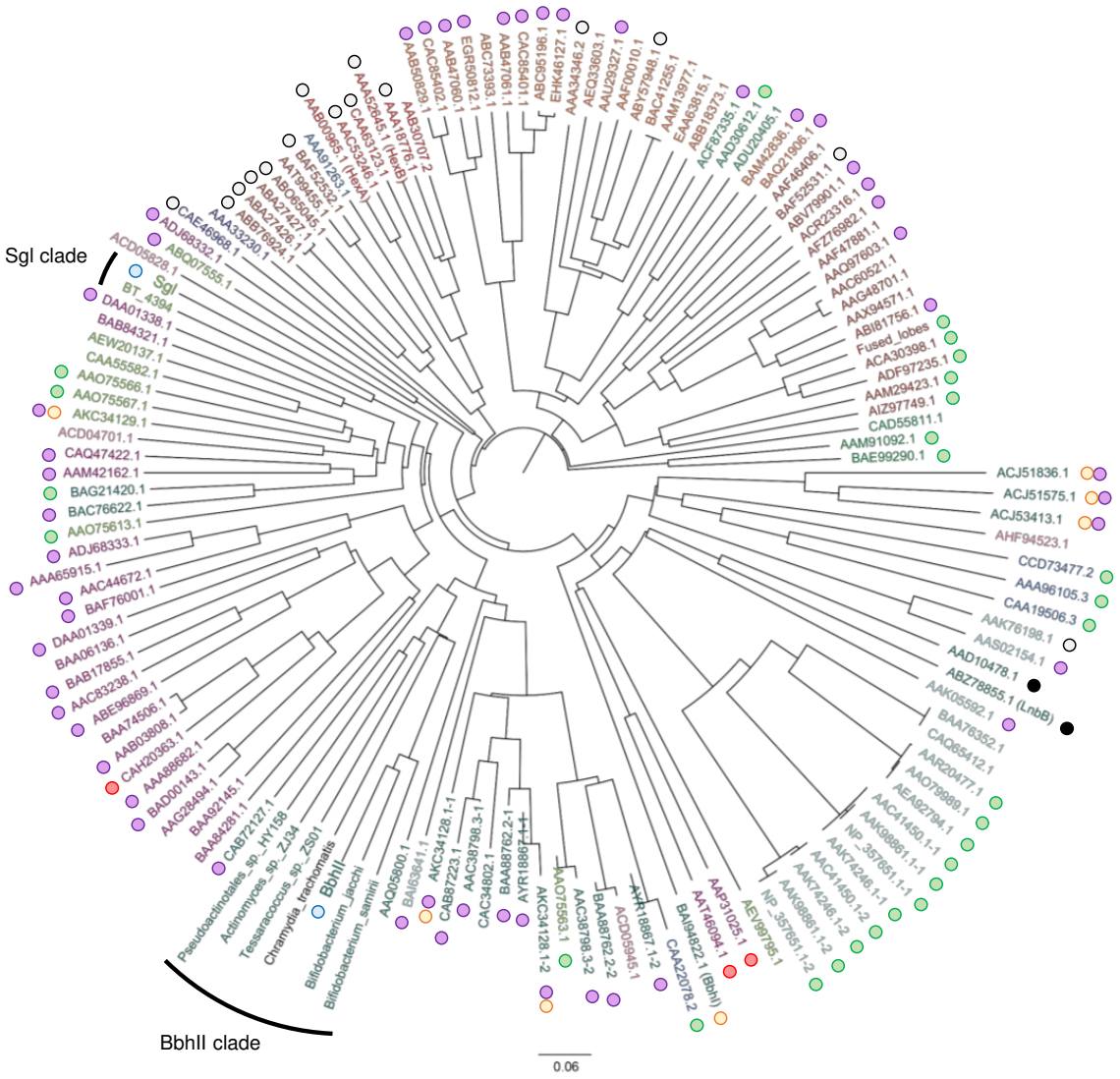
Katoh *et al.*, Extended Data Fig. 1







Katoh *et al.*, Extended Data Fig. 2



Substrate specificity

- 6-SO_3-β-*N*-Acetylglucosaminidase
- β-*N*-Acetylhexosaminidase (Broad substrate specificity)
- Chitinolytic β-*N*-acetylglucosaminidase
- *N*-Glycan processing β-*N*-acetyglucosaminidase
- Disperisin B (β-1,6-specific *N*-acetylglucosaminidase)
- β-1,3-*N*-Acetylglucosaminidase acting on lacto-*N*-triose II and mucin core 3
- Lacto-*N*-biosidase

Taxonomic classification

Eukaryotes
- Mammal
- Plant
- Insect/prawn
- Fungi
- Nematode (*C. elegans*)
- Others

Bacterial phyla
- Actinomycetota
- Bacillota
- Bacteroidota
- Chlamydiae
- Pseudomonadota
- Verrucomicrobiota

Katoh *et al.*, Extended Data Fig. 3

a Relative abundances of species-specific reads vs. *bbhII* homolog-specific reads in the metagenomic dataset obtained for mothers

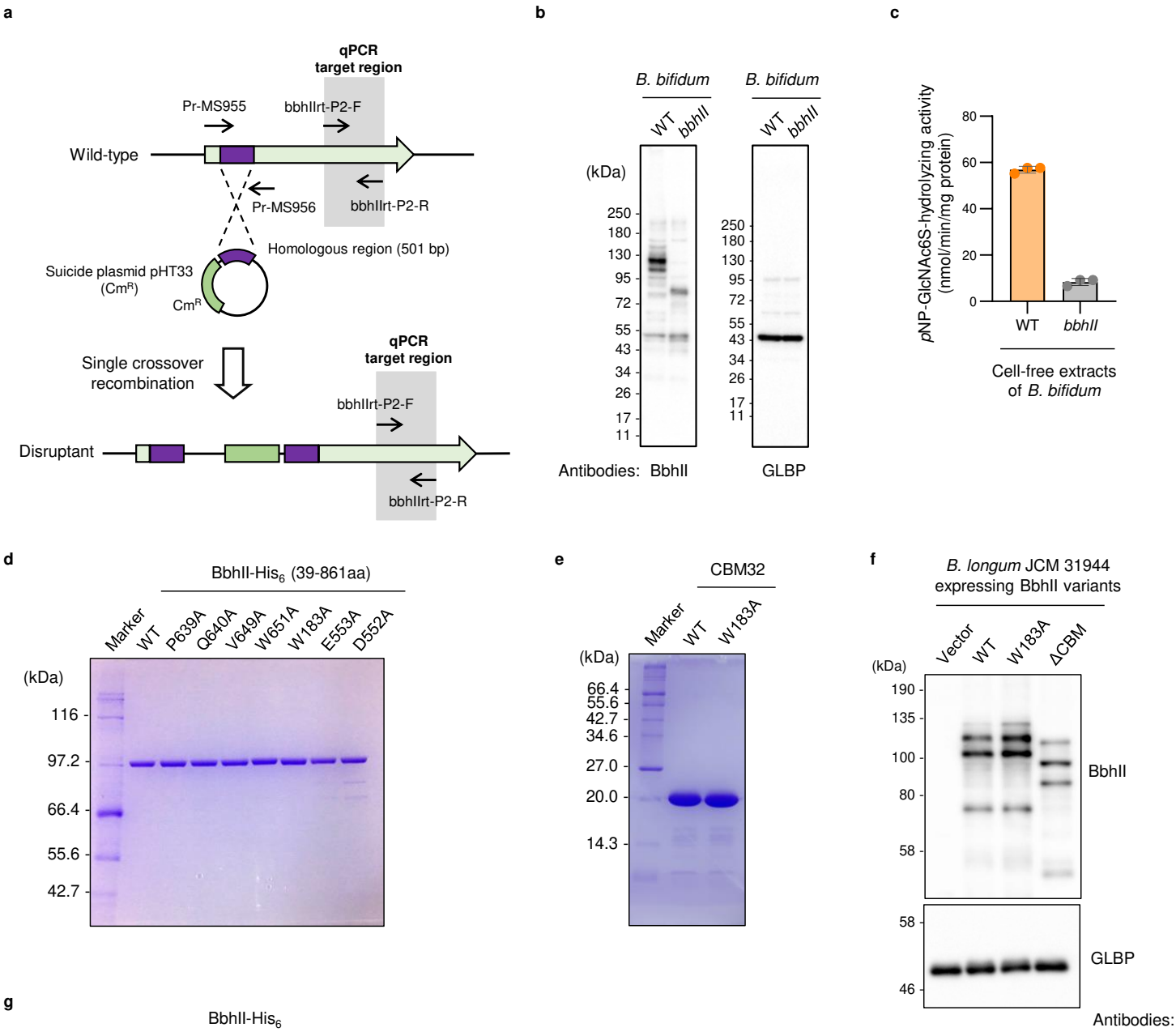
Species	q value (FDR correction, Q = 5 %)	r
Flintibacter sp. KGMB00164	0.0167	0.378

b Relative abundances of species-specific reads vs. *sgl* homolog-specific reads in the metagenomic dataset obtained for mothers

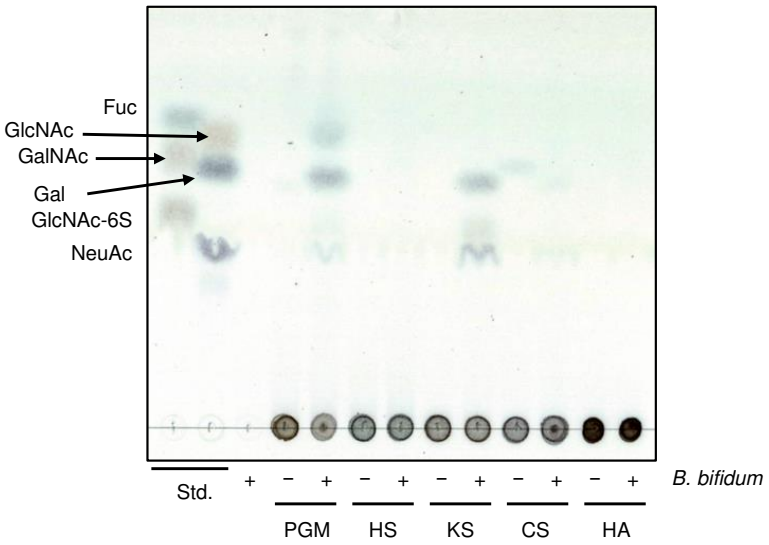
Species	q value (FDR correction, Q = 5 %)	r
Bacteroides thetaiotaomicron	0	0.798
Bacteroides caccae	0.0000437	0.5228
Parabacteroides sp. CT06	0.00313	0.423
Bacteroides sp. HF 162	0.00313	0.418
Bacteroides caecimuris	0.00313	0.416
Bacteroides sp. HF 5141	0.00313	0.411
Bacteroides sp. CBA7301	0.00313	0.405
Bacteroides fragilis	0.00313	0.405
Bacteroides sp. CACC 737	0.00393	0.396
Bacteroides ovatus	0.00647	0.378
Bacteroides uniformis	0.0127	0.356
Bacteroides xylanisolvens	0.0152	0.348
Parabacteroides distasonis	0.0152	0.346
Bacteroides sp. A1C1	0.0232	0.331
Lacrimispora saccharolytica	0.0294	-0.321
Bifidobacterium angulatum	0.0300	-0.319
Bacteroides sp. M10	0.0483	0.301

c Relative abundances of species-specific reads vs. *sgl* homolog-specific reads in the metagenomic dataset obtained for infants

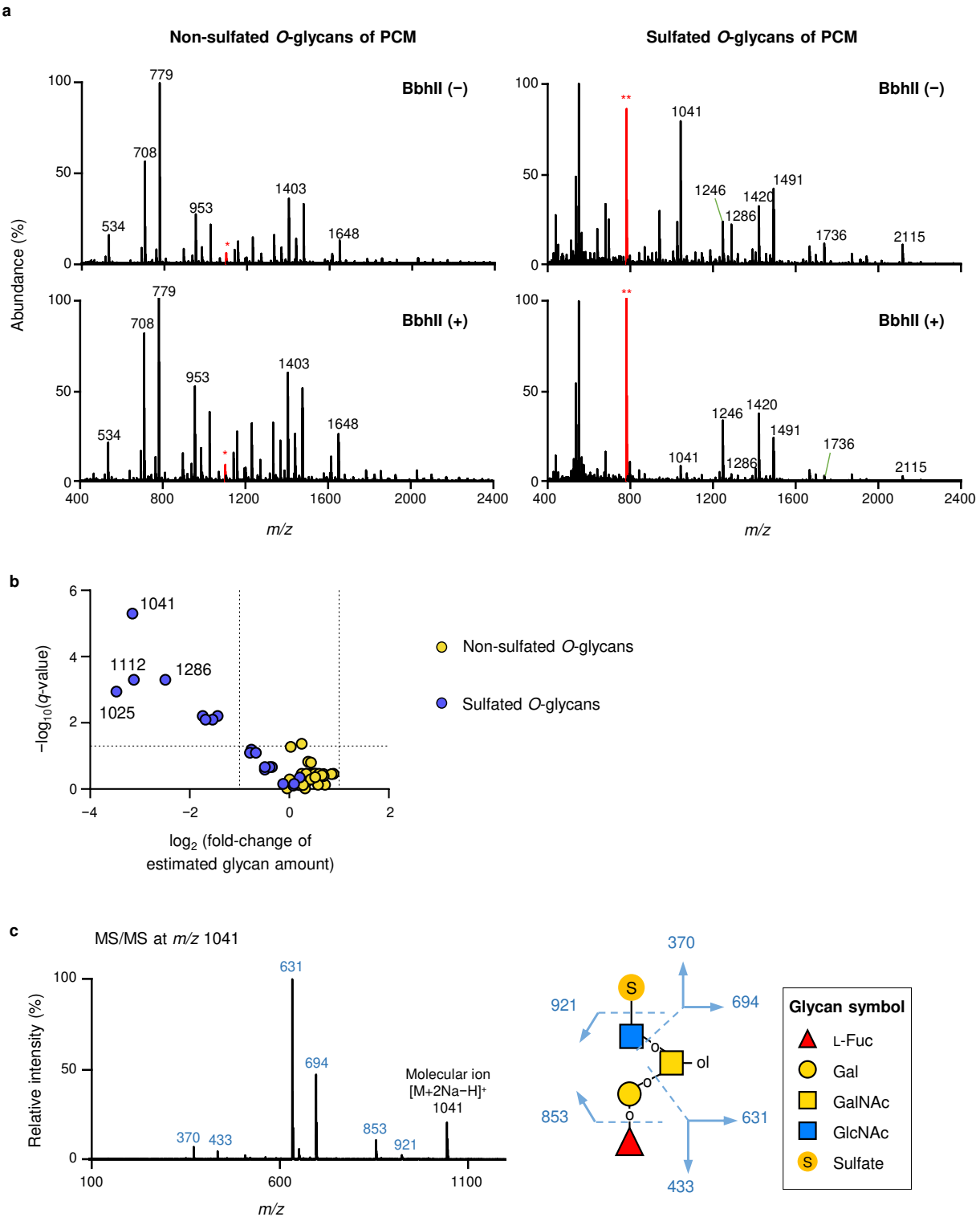
Species	q value (FDR correction, Q = 5 %)	r
Bacteroides thetaiotaomicron	2.28E-14	0.764
Bacteroides caccae	3.43E-13	0.737
Bacteroides sp. CBA7301	3.84E-08	0.617
Butyricimonas faecalis	1.57E-07	0.595
Bacteroides sp. A1C1	4.23E-07	0.573
Bacteroides sp. HF 162	4.23E-07	0.572
Bacteroides sp. CACC 737	4.23E-07	0.572
Paraprevotella xylaniphila	2.21E-06	0.545
Barnesiella viscericola	1.29E-05	0.513
Bacteroides uniformis	1.29E-05	0.511
Bacteroides intestinalis	1.78E-05	0.504
Phocaeicola dorei	1.78E-05	0.502
Bacteroides sp. M10	2.45E-05	0.495
Odoribacter splanchnicus	2.93E-05	0.490
Bacteroides sp. HF 5287	3.14E-05	0.487
Bacteroides helcogenes	3.95E-05	0.481
Bacteroides caecimuris	0.0000500	0.476
Bacteroides cellulosilyticus	7.80E-05	0.465
Bacteroides sp. HF 5141	7.80E-05	0.464
Parabacteroides sp. CT06	0.000213	0.442
Bacteroides ovatus	0.000262	0.436
Parabacteroides distasonis	0.000416	0.424
Phocaeicola salanitronis	0.000461	0.421
Alistipes megaguti	0.00112	0.397
Collinsella aerofaciens	0.00132	0.392
Clostridium perfringens	0.00207	-0.379
Bacteroides xylanisolvens	0.00337	0.365
Phocaeicola vulgatus	0.00349	0.363
Sutterella faecalis	0.00425	0.356
Parolsenella catena	0.00754	0.338
Alistipes finegoldii	0.00923	0.330
Alistipes shahii	0.00983	0.327
Alistipes communis	0.0156	0.311



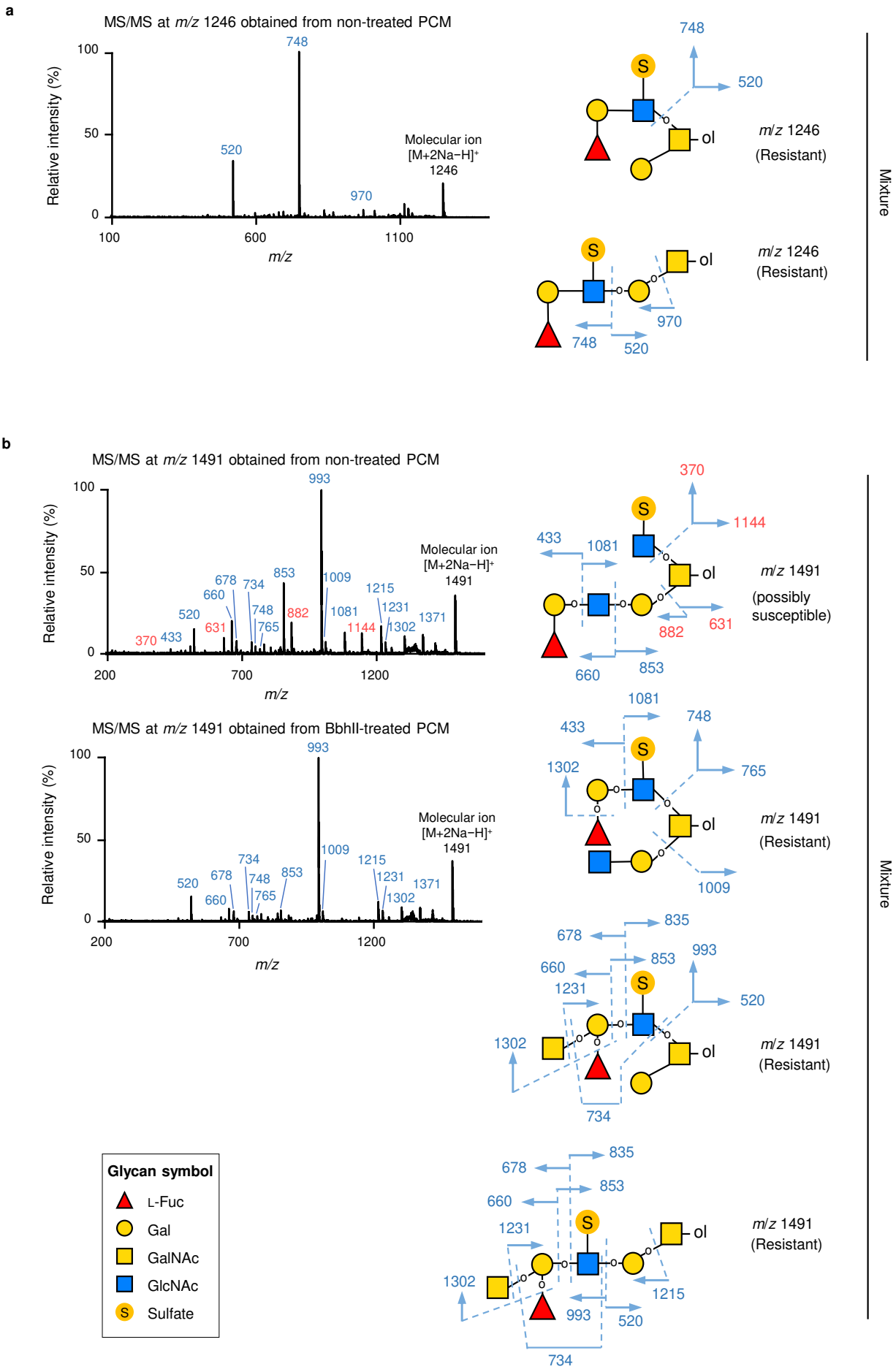
Katoh *et al.*, Extended Data Fig. 5



Katoh *et al.*, Extended Data Fig. 6

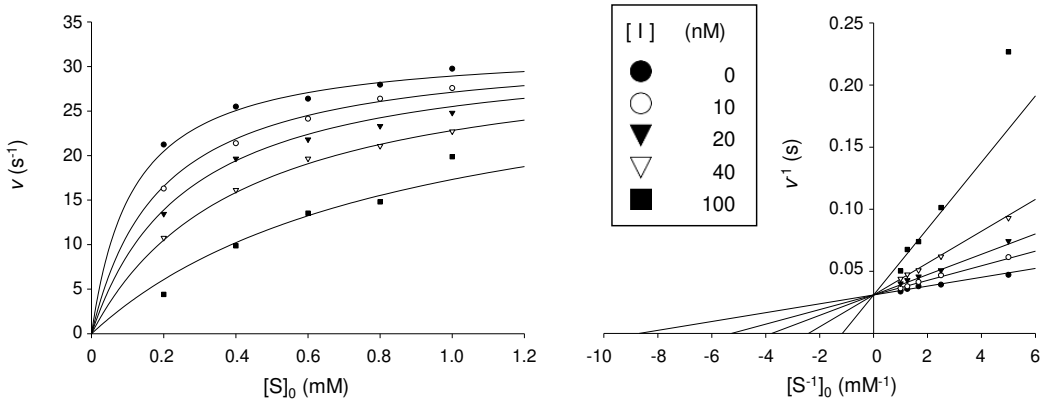


Katoh *et al.*, Extended Data Fig. 7

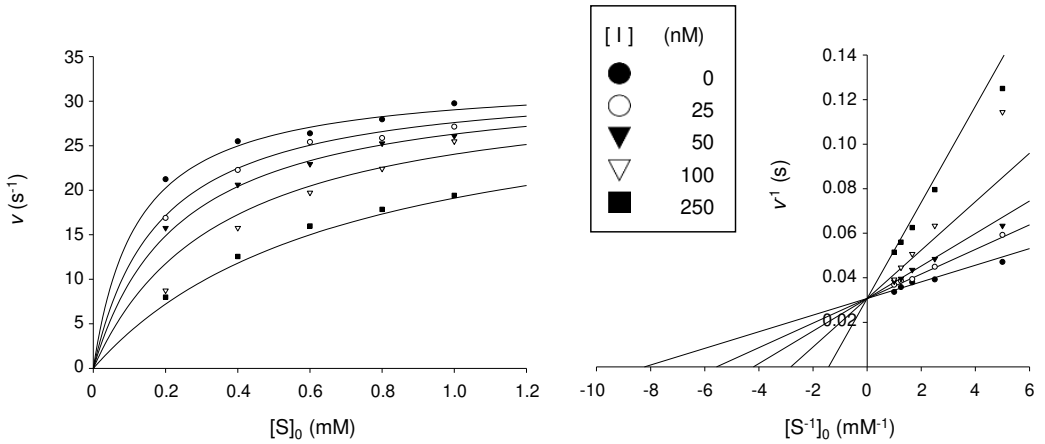


Katoh *et al.*, Extended Data Fig. 8

a Inhibition of BbhII by PUGNAc-6S



b Inhibition of BbhII by NAGT-6S

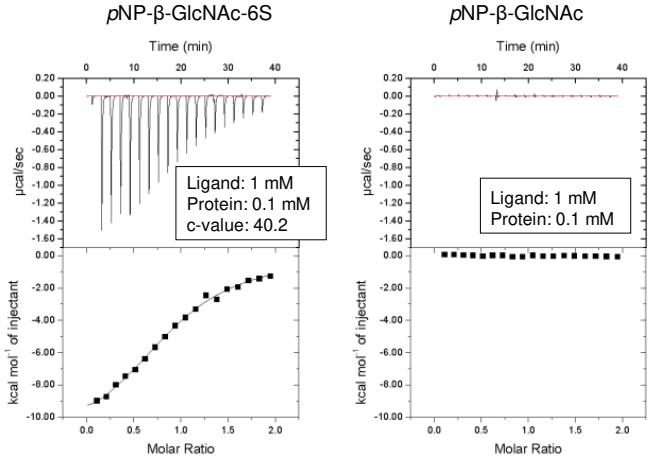


Inhibitor	Mode	K_m (μM)	k_{cat} (s^{-1})	K_i (nM)
PUGNAc-6S	Competitive	115 ± 14	32.3 ± 0.7	15.4 ± 1.9
NAGT-6S	Competitive	122 ± 16	32.6 ± 0.8	52.3 ± 7.2

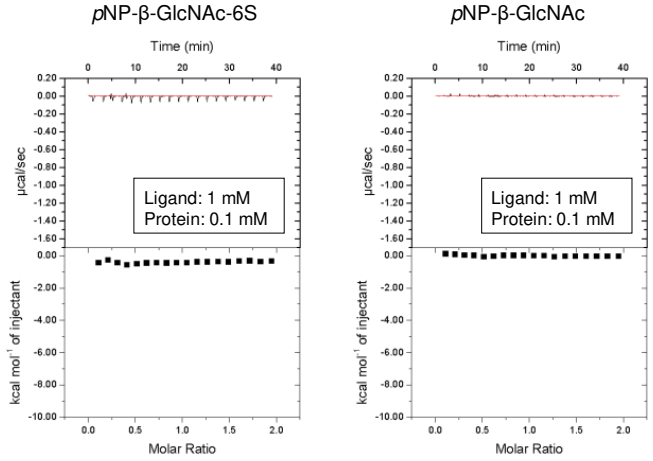
Equation

$$v = \frac{k_{cat}\,[E]_0\,[S]}{K_m\,(1 + [I] / K_i) + [S]}$$

c ITC analysis of CBM-His_6 (WT)

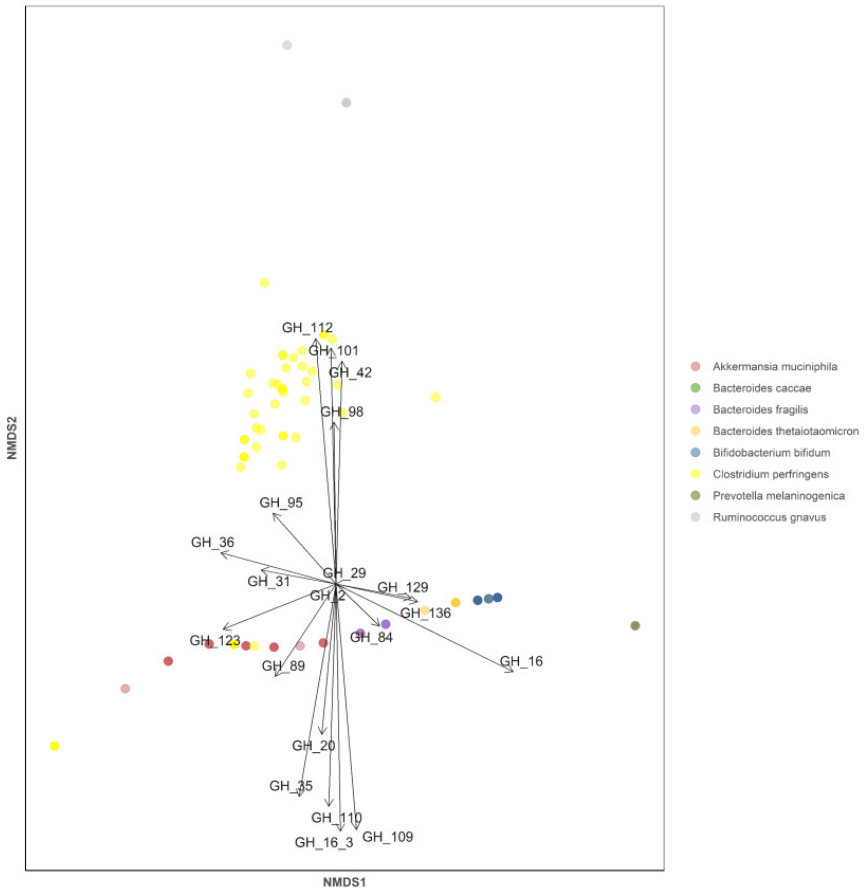


d ITC analysis of CBM-His_6 (W183A)



	K_a (10^3 M^{-1})	K_d (μM)	ΔG^0 (kJ mol^{-1})	ΔH (kJ mol^{-1})	$-T\Delta S^0$ (kJ mol^{-1})	n
pNP-β-GlcNAcc-6S	40.2 ± 3.4	24.9 ± 2.1	−26.7	−49.2 ± 1.5	22.5	0.924 ± 0.018

Katoh *et al.*, Extended Data Fig. 9



VECTORS

muc-GH	NMDS1	NMDS2	R^2	P (>r)
GH16_3	0.01789	-0.99984	0.5380	0.0001
GH109	0.08254	-0.99659	0.5344	0.0001
GH112	-0.08254	0.99659	0.5344	0.0001
GH101	-0.02202	0.99976	0.4907	0.0001
GH42	0.02708	0.99963	0.4381	0.0001
GH110	-0.03276	-0.99946	0.4358	0.0001
GH35	-0.17121	-0.98523	0.4095	0.0001
GH16	0.89625	-0.44354	0.3438	0.0001
GH98	-0.01299	0.99992	0.2307	0.0001
GH20	-0.09459	-0.99552	0.2009	0.0001
GH123	-0.92867	-0.37091	0.1303	0.0001
GH36	-0.96505	0.26208	0.1253	0.0003
GH89	-0.55027	-0.83499	0.1076	0.0001
GH95	-0.66414	0.74761	0.0793	0.0002
GH136	0.97722	-0.21225	0.0612	0.0028
GH31	-0.98287	0.1843	0.0514	0.0092
GH129	0.98287	-0.1843	0.0514	0.0092
GH84	0.71886	-0.69515	0.0322	0.0397
GH2	0	0	0	1
GH29	0	0	0	1

Permutation: free

Number of permutations: 9999